Title
Eye gaze patterns reveal how we reason about fractions

Journal
Thinking & Reasoning, 24(4)

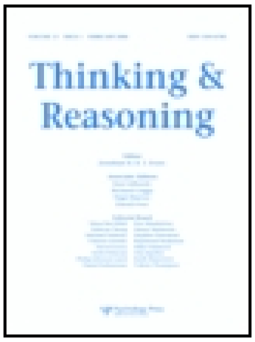
ISSN
1354-6783

Authors
Singley, Alison T Miller
Bunge, Silvia A

Publication Date
2018-10-02

DOI
10.1080/13546783.2017.1417909

Peer reviewed



Routledge
Taylor & Francis Group

Thinking & Reasoning

ISSN: 1354-6783 (Print) 1464-0708 (Online) Journal homepage: http://www.tandfonline.com/loi/ptar20

Eye gaze patterns reveal how we reason about fractions

Alison T. Miller Singley & Silvia A. Bunge

THINKING & REASONING, 2017
https://doi.org/10.1080/13546783.2017.1417909

Eye gaze patterns reveal how we reason about fractions

Alison T. Miller Singley[a] and Silvia A. Bunge[a,b]

[a]Department of Psychology, University of California at Berkeley, Berkeley, CA, USA; [b]Helen Wills Neuroscience Institute, University of California at Berkeley, Berkeley, CA, USA

ABSTRACT
Fractions are defined by numerical relationships, and comparing two fractions' magnitudes requires within-fraction (holistic) and/or between-fraction (componential) relational comparisons. To better understand how individuals spontaneously reason about fractions, we collected eye-tracking data while they performed a fraction comparison task with conditions that promoted or obstructed different types of comparisons. We found evidence for both componential and holistic processing in this mixed-pairs task, consistent with the hybrid theory of fraction representation. Additionally, making within-fraction eye movements on trials that promoted a between-fraction comparison strategy was associated with slower responses. Finally, participants who performed better on a non-numerical test of reasoning took longer to respond to the most difficult fraction trials, which suggests that those who had greater facility with non-numerical reasoning attended more to numerical relationships. These findings extend prior research and support the continued investigation into the mechanistic links between numerical and non-numerical reasoning.

ARTICLE HISTORY Received 16 February 2017; Accepted 11 December 2017

KEYWORDS Eye-tracking; mathematical reasoning; relational reasoning; fractions; magnitude

Introduction

Reasoning and math

Everyday issues, both important and trivial, require that we reason about mathematical relationships. *The United States government spends 8.2% of GDP on health care, and the private sector spends an additional 9.8%; the Canadian government spends 11% of GDP on health care, so which system is the better deal? Generic facial tissue is half the price of Kleenex, but when there's a buy-2-get-1-free sale on Kleenex, perhaps the little extra splurge is worth it?*

CONTACT Alison T. Miller Singley atms@berkeley.edu
Supplemental data for this article can be accessed here https://doi.org/10.1080/13546783.2017.1417909.

That reasoning supports mathematical thinking is also apparent in laboratory settings: tests of specific mathematical skills strongly correlate with tests of domain-general relational reasoning skills (McGrew & Hessler, 1995; Morsanyi, Devine, Nobes, & Szűcs, 2013), and in some cases, relational reasoning performance reflects mathematical skill (Prado, Van der Henst, & Noveck, 2008). Moreover, there is an emerging longitudinal link between the two, such that performance on reasoning tests predicts future mathematical skill (Green, Bunge, Briones Chiongbian, Barrow, & Ferrer, 2017; Primi, Ferrão, & Almeida, 2010), and current educational policy promotes the use of analogical skills for conceptual development in mathematics learning (Richland & Begolli, 2016). However, the mechanistic links between relational reasoning and mathematical cognition are not yet understood.

Focus on fractions

Fractions provide an ideal testing ground in which to explore the intersection of mathematics and reasoning. As one way to represent rational numbers, they are inherently a mathematical concept – and yet, because they are defined by a numerical relationship, we contend that understanding fractions' magnitudes also relies on domain-general relational reasoning. Two prevailing theories about the nature of the mental representation of fractions capture the dual aspects of their definition (Bonato, Fabbri, Umilta, & Zorzi, 2007; Meert, Grégoire, & Noël, 2009). The holistic theory holds that fractions are represented by their integrated magnitude. Under this processing model, initial encoding of a fraction (e.g., 2/3) requires a mathematical calculation or estimation, and subsequent processing manipulates only the resulting magnitude (e.g., 0.67), making operations like addition or comparison of two fractional values relatively straightforward. By contrast, the componential theory states that the two component numbers, the numerator and denominator, are mentally held as separate entities. Per the componential processing model, the fraction must be encoded as a relation between two components (e.g., 2 and 3), and all subsequent processing must take into account the relationship between numerator and denominator, and, in the case of operations, the relationships between components of different fractions (e.g., addition requires the same denominators). Thus, the componential model presupposes relational reasoning to a greater extent than does the holistic model.

These competing mental models predict distinct behavioural signatures in a fraction comparison task, in which participants are asked to select the greater of two fractions from a display. This paradigm traditionally makes use of the distance effect, which is the well-documented phenomenon that people take longer to distinguish between two numbers that are closer together than those that are farther apart (Moyer & Landauer, 1967). For example, it is easier to judge 2/11 vs. 8/9 than it is to judge 3/5 vs. 4/7, because both the numerators and the magnitudes of the first fraction pair are relatively far apart. Given this phenomenon, manipulating the distance between the components versus the distance between the fraction magnitudes can provide an indication of whether participants mentally represent the components or the magnitudes of a fraction. This methodology has provided empirical evidence for both componential processing (Bonato et al., 2007; Ischebeck, Schocke, & Delazer, 2009) and holistic magnitude processing (Ischebeck et al., 2009; Sprute & Temple, 2011). Additional studies that have shown behaviour indicative of both componential and holistic processing depending on the task context (Faulkenberry & Pierce, 2011; Fazio, DeWolf, & Siegler, 2016; Gabriel, Szucs, & Content, 2013; Meert et al., 2009; Meert, Grégoire, & Noël, 2010; Smith, 1995) have given rise to a third theory that both components and magnitudes are accessible via a hybrid mental representation (Meert et al., 2009).

However, the particular stimuli presented may promote a certain strategy. For example, when two fractions to be compared share a common component, e.g., 2/5 vs. 2/7, it encourages componential processing because only the non-identical numbers need to be compared. Indeed, much of the evidence supporting the componential theory comes from studies that use shared-component stimuli (Bonato et al., 2007; Ischebeck et al., 2009). Evidence for the holistic theory has primarily come from studies using stimuli without shared components (Faulkenberry, Montgomery, & Tennes, 2015; Gabriel et al., 2013; Smith, 1995; Sprute & Temple, 2011; Zamarian, Ischebeck, & Delazer, 2009), suggesting that people attend to holistic magnitudes when a simple comparison between two components is not available. Additionally, using over-learned or familiar fractions, such as 7/10 vs. 3/4, makes it easier to assess the holistic magnitude of the fraction and thus may promote holistic processing. The finding that people are sensitive to these manipulations and adjust their approach accordingly constitutes strong evidence for the hybrid representation model. However, it is an open question whether both componential and holistic representations are beneficial beyond the specific cases noted above.

Because relational and mathematical reasoning are closely linked, it is also informative to characterise these prior findings in the context of relational reasoning. Halford and Wilson first proposed a method of characterising cognitive development by examining the number of cognitive elements and complexity of relationships between elements required to mentally represent a task (Halford & Wilson, 1980). This theory was subsequently developed into a system for measuring relational complexity (Halford, Wilson, & Phillips, 1998).

In fraction pairs with shared components, only the non-shared numbers are relevant, and thus only a simple comparison between two numbers is necessary to determine which fraction has the larger magnitude. In this case, the componential processing model provides the simplest relational comparison. However, when fraction pairs do not share components, the holistic model provides for simpler relational processing. To apply terminology from relational reasoning tasks, estimating the fraction magnitude is a first-order, or within-fraction, comparison (Miller Singley & Bunge, 2014). Because reducing that first-order relationship to a single number reduces its complexity (English & Halford, 1995), the subsequent comparison between two magnitudes is another simple first-order comparison. For example, the pair 2/3 versus 5/12 could be solved by estimating that 5/12 is less than 0.5 while 2/3 is greater.

Componential processing of mixed pairs requires several between-fraction evaluations: it is first necessary to compare the numerators to identify the multiplicative relationship between them, then the denominators (or vice versa), and then compare those two multiplicative relationships to each other. Using the terminology of relational reasoning, this is a second-order comparison of two first-order relationships (Miller Singley & Bunge, 2014). Approaching the same example of 2/3 versus 5/12 with a componential strategy involves noting that the second denominator (12) is 4 times the first (3) while the second numerator (5) is only 2.5 times the first (2). Thus, approaching a mixed-pair evaluation using componential processing is more onerous from a relational reasoning perspective than using holistic processing. Therefore, the prior findings – that people use componential processing when a simple comparison is sufficient and holistic processing when all four numbers must be evaluated – align with relational theory.

In summary, the case for a hybrid mental representation of fractions largely rests on the experimental stimuli presented, with shared components tasks yielding componential processing and mixed-pairs tasks showing evidence of holistic processing, a switch that aligns with relational complexity theory. However, a shared components context is a special case of the fraction comparison task, and it is not yet known whether the hybrid model applies in a more general mixed-pairs context.

To probe the hybrid theory, we tested for the use of componential and holistic processing in a mixed-pairs fraction comparison task in which some conditions promoted holistic processing and some promoted componential. Relational complexity theory suggests that holistic processing is always simpler in a mixed-pairs context, and yet hybrid theory suggests people may be sensitive to task manipulations that privilege one model over another. In accordance with hybrid theory, we predicted that people would use processing models that align to the affordances of the task conditions.

Furthermore, directly comparing numerical and non-numerical relational reasoning tasks may illuminate the mechanisms that underpin both skills. Non-numerical reasoning tasks can be characterised by first- and second-order comparisons; we selected the Analysis-Synthesis sub-test of the Woodcock–Johnson III Tests of Cognitive Abilities (Woodcock, McGrew, & Mather, 2001), which involves a given set of first-order relationships that must be combined across several second-order relationships to ascertain correct responses. We predicted that numerical and non-numerical reasoning would be correlated on these two tasks.

Probing fraction comparison strategies via eye-tracking

Chronometric studies, such as those described above, enable inferences about the mental representation of fractions; however, cognitive processing can be more directly observed using eye-tracking methodology. It is well established that eye movements reflect the focus of attention on a moment-to-moment basis (Shepherd, Findlay, & Hockey, 1986), and that the duration and trajectory of eye movements reflect underlying cognitive processes (Just & Carpenter, 1976). Applying eye-tracking technology in the context of a fraction comparison task generates metrics on eye movements between digits as an indication of which numbers people are actively comparing. Eye-tracking therefore provides insights as to whether people are using componential or holistic processing as they compare fraction magnitudes.

The few relevant eye-tracking studies to date have supported the hybrid theory of fraction representation, consistent with the bulk of the behavioural studies. Eye movements, or saccades, between components of different fractions, such as from one numerator to the other, are taken as evidence of componential processing; saccades between a numerator and denominator of the same fraction are taken as evidence of holistic processing. Ischebeck, Weilharter, and Körner (2016) compared trials in which the fraction pairs shared a common component, such as 3/5 vs. 4/5, to mixed pairs, such as 2/5 vs. 3/7. They found that people exhibit more saccades between the non-identical components than between the shared components on common components trials (e.g., more saccades between 3 and 4 when comparing 3/5 to 4/5 than between the two 5s). Obersteiner and Tumpek (2016) reported similar results, and additionally found that people exhibited behaviour more consistent with magnitude processing (i.e., more numerator–denominator saccades) when viewing mixed pairs than pairs with shared components.

As with the prior behavioural experiments, these studies support the hybrid model that people adaptively switch between mental representations depending on task demands, and do so in accordance with relational complexity theory, yet the evidence is largely based on the presence of the specific case of shared components. To investigate whether hybrid theory applies in the general case of mixed-pair comparisons, we tracked saccades during more subtle manipulations to investigate whether and to what extent people use componential versus holistic processing.

Research questions and hypotheses

In the present study, we investigated the type of processing employed during performance of a difficult version of the fraction comparison task, and the resulting impact on performance. We selected two manipulations that should promote opposite processing approaches – between-fraction/componential or within-fraction/holistic comparisons – and tested the participants' eye movements and performance. The two manipulations were (1) making one denominator a multiple of the other, which promotes between-fraction or componential processing, and (2) placing the magnitude of the fractions on opposite sides of 1/2, which promotes within-fraction or holistic processing. These manipulations align with commonly taught or commonly used strategies for evaluating fractions: converting to equivalent fractions, and benchmarking to a common referent (Clarke & Roche, 2009; Smith, 1995). Following Ischebeck et al. (2016) and Obersteiner and Tumpek (2016), saccades within fractions were taken as evidence for holistic processing; saccades between fractions were taken as componential processing. We tracked the number of each type of saccade to investigate the extent to which the different types of processing contributed to performance.

Per hybrid theory, we expected to see gaze behaviour indicative of both componential and holistic processing throughout the task. If participants engaged more readily in componential processing, we would predict better performance on trials that promoted componential processing (the trials in which one denominator is a multiple of the other). If, on the other hand, they engaged more readily in holistic processing, we would predict better performance on trials that facilitated the strategy of benchmarking by magnitude.

With respect to eye gaze, we hypothesised that participants would exhibit more horizontal saccades on trials with multiples and more vertical saccades on trials with magnitudes opposite 1/2, relative to the individuals' overall average behaviour. Furthermore, we hypothesised that condition-aligned gaze behaviour would be associated with better performance, such that the extent to which the gaze behaviour reflected the type of processing promoted by the trial condition would predict performance on that trial. Finally, we tested whether performance and the proportion of horizontal saccades on this task would be related to performance on a visuospatial relational reasoning task, with the hypothesis that better relational reasoning would be associated with better performance on the fractions task and more componential processing.

Materials and methods

Participants

Thirty-eight participants from the university research subject pool completed one-hour-long session for course credit in the Department of Psychology. The participants' ages ranged from 18 to 22 (Mean = 20.3, SD = 1.2) and the group was ethnically diverse, reflecting the Bay Area population. Of these participants, 25 self-identified as female, 9 as male, and 4 declined to state their gender. We recruited participants at the university level to ensure familiarity with fractions and arithmetic. All participants had completed at least one semester of university-level mathematics or statistics, and 74% rated themselves as "somewhat" or "very" confident in their mathematics skills (77% of final sample). Study procedures were approved by the Committee for the Protection of Human Subjects at the University of California, Berkeley.

The one-hour testing session included five blocks of a fraction comparison task. Participants first reviewed and signed the consent form and completed a brief demographic survey asking for their major, math experience, math confidence, and ethnicity. Block 1, which was self-paced and lasted 2–5 minutes, was a variation of this fraction comparison task, adapted from Ischebeck et al. (2009). It included fraction pairs with the same numerator or the same denominator, along with pairs that shared no common components. This first block was acquired as a point of comparison for a developmental study reported elsewhere (Crawford, Miller Singley, & Bunge, in preparation). Blocks 2 and 3, each lasting approximately 3–5 minutes, formed the basis of the present investigation.

To provide a break from the computerised eye-tracking task, the blocks were separated by the Analysis-Synthesis sub-test of the Woodcock–Johnson III Tests of Cognitive Abilities (Woodcock et al., 2001). This test was administered as an independent measure of domain-general relational reasoning. After completing Block 3, participants completed two final variations of the fraction task, not reported here: Block 4 used improper and proper fractions in which each pair had either the same numerators or same denominators, and Block 5 contained one sample trial from each task condition seen throughout all runs, during which the participants self-reported the strategies they used to solve each problem.

We set a criterion of 60% valid eye-tracking data (e.g., Kafkas & Montaldi, 2012) for the two blocks included in this study; two participants did not meet this criterion and were excluded from further analyses. Five additional participants' performance did not differ from chance (50% correct overall; based on one-sample probability tests, p ranged from 0.07 to 0.89), so they were also excluded. For the gaze analyses, we evaluated only the trials that included specific saccades of interest: those between the four numbers on the screen. After excluding trials without saccades of interest, one participant's average accuracy on remaining trials was three standard deviations below the mean accuracy of the sample, and this participant was retroactively dropped from all analyses. Thus, the analyses described in this study were conducted on data from 30 participants.

Study design

On each trial, participants were asked to select the larger of two fractions in a given pair, within a 4-second time window. None of the fraction pairs used in

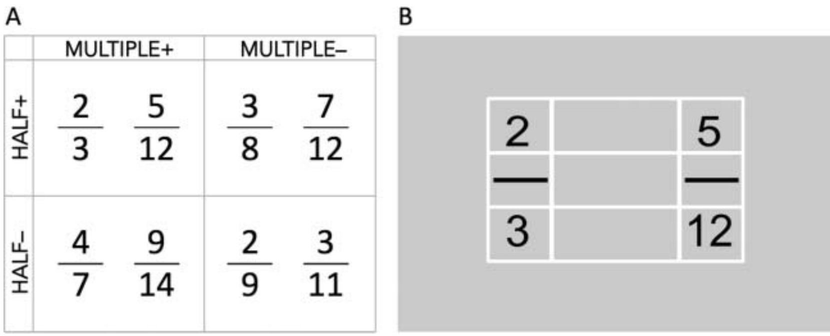


Figure 1. (A) Two-factorial task design. In half of the trials, one denominator was a multiple of the other (M+). In half of the trials, the fraction magnitudes spanned 0.5 (H+). (B) Areas of interest displayed on a sample screen layout.

the comparison task shared common components, similar to the more difficult conditions from prior studies (Ischebeck et al., 2016; Obersteiner & Tumpek, 2016). Trials followed a 2 × 2 factorial experimental design, described below (Figure 1, panel A). We selected two common problem-solving strategies, one involving componential processing and the other reflecting holistic processing, and created fraction pairs that we reasoned should promote one of these strategies over the other. We did not mention either strategy to subjects or ask them to approach the problems in a particular way; rather, we used patterns of eye gaze as an indicator of mental representations of fractions.

The four conditions described below were interspersed pseudorandomly throughout the two task blocks. The first three participants were given 10 s to complete each trial, but we found they completed the trials much more quickly and so we adjusted the time limit to 4 s for the remaining participants. Trials on which participants did not respond within the 4-second timeframe were marked as incorrect, and for the first three participants, we re-coded as incorrect any of their correct responses that were made after 4 s.

The Multiple factor governed the denominators of the two fractions in a pair, wherein half of the pairs had one denominator that is a multiple of the other (e.g., 4/7 vs. 5/14), referred to subsequently as M+, whereas the other half did not (e.g., 4/7 vs. 6/13; referred to as M−). Some M+ trials also had numerators that were multiples of each other, but none of the M− trials did. M+ trials were designed to promote the use of the equivalent fractions strategy, in which one fraction is converted to an equivalent fraction by multiplying both numerator and denominator by the same factor. In the given example, 4/7 × 2/2 = 8/14, which is an easy comparison to 5/14. In M+ trials, only one fraction needed to be converted, and the multiple was easily identified. We consider this strategy to reflect componential processing, as the focus of attention is on the relationships between the numerators and denominators, and ultimately requires a second-order comparison between those two relationships (e.g., “double” vs. “more than double”). Attempting to use the same strategy on M− trials, such as 4/7 vs. 6/13, would require both fractions to be converted by the least common multiple; this is both mentally challenging and extremely difficult to complete within the trial time limit. It would be more efficient to use a different strategy, such as benchmarking to 1/2, described below, or cross-multiplying, which involves multiplying each numerator by the opposite fraction's denominator and comparing the cross-products.

The Half factor governed the magnitude of the fractions, such that half of the pairs had one fraction whose magnitude was less than 1/2 and the other fraction was greater than 1/2 (e.g., 4/7 vs. 6/13), referred to subsequently as H+ trials. H+ trials were designed to promote the benchmarking to 1/2 strategy, in which one can quickly estimate or calculate whether each fraction's magnitude is greater or less than 1/2, and then make a simple numerical comparison between the two magnitudes. We consider this strategy to reflect holistic processing because the focus of attention is on the relationship between the numerator and denominator that defines the magnitude of each fraction. For H− trials, wherein the fraction magnitudes are both greater than or both less than 1/2 (e.g., 4/7 vs. 9/14), benchmarking to 1/2 is not helpful. It would be more efficient to choose a different strategy, such as finding an equivalent fraction or cross-multiplying.

Crossing these two factors led to four conditions: M+H+, M+H−, M−H+, M−H− (Figure 1, panel A). The first condition was designed to facilitate the use of either componential or holistic processing, because it promoted both strategies of converting to equivalent fractions and benchmarking to ½, and we predicted that participants would perform the best on this condition. The M+H− condition was designed to facilitate componential processing, and therefore we expected that participants would make more horizontal than vertical saccades on these trials, as they focused on comparing components. The M−H+ condition was designed to facilitate holistic processing, and therefore we expected that participants would make more vertical than horizontal saccades as they estimated the magnitude of each fraction. An initial scan is necessary to discern the trial type, and Ischebeck and colleagues found the initial scan pattern to be somewhat idiosyncratic (Ischebeck et al., 2016). Thus, we expected each trial to include both horizontal and vertical saccades, but we also expected that the majority of saccades would reflect the type of processing best suited to the trial's condition.

Finally, the M−H− condition obstructed both strategies, so we expected worse performance on these trials than on the others. We considered the possibility that participants would resort to a cross-multiplication strategy on M−H− trials, which would be reflected in predominantly diagonal saccades.

In fact, however, the results do not bear this out, as diagonal saccades made up only 5% of all recorded saccades, and there were no more diagonal saccades on M−H− trials than on others. Below, we consider only horizontal and vertical saccades as indices of componential and holistic processing, respectively.

Prior research on componential strategies has shown that, when a simple comparison between either numerators or denominators is sufficient to solve the comparison problem, people only make that single comparison (Bonato et al., 2007; Ischebeck et al., 2009). Therefore, following Ischebeck's "incongruent" condition, we used only pairs in which one fraction had both a larger numerator and a larger denominator than the other fraction. In such cases, the numerators and denominators provide conflicting cues about the magnitudes of fractions. For example, when judging the pair 3/5 vs. 4/7, the larger numerator (4) implies a larger fraction and the larger denominator (7) implies a smaller fraction. Thus, participants had to take all four numbers into account to perform these trials correctly.

Prior research also indicates that when the magnitude difference between the two fractions is greater than 0.3, adults' performance on the comparison task approaches ceiling (Gabriel et al., 2013; Sprute & Temple, 2011). Thus, we ensured that the magnitude difference between all fractions was less than 0.3; actual values ranged from 0.01 to 0.27. Within this restricted range, we still saw wide variability in accuracy: average trial accuracy ranged from 28% of participants answering correctly to 100% (Figure 2, panel C).

Furthermore, given the documented whole number bias (Ni & Zhou, 2005), it is tempting to assume that larger numbers comprise the larger fraction, yet 2/3 represents a greater magnitude than 12/19. In half of the trials, the larger numbers indeed comprised the larger fraction; in half they did not. Finally, all fractions were proper (between 0 and 1), irreducible, composed of digits 1–19, and the stimuli were counterbalanced such that half of the correct answers were on the left and half on the right. The full list of stimuli is provided in Online Supplementary Table S1.

Data collection

The experiment was conducted on a Tobii T120 eye-tracker, with a sampling rate of 120 Hz (or every 8.3 ms). Participants sat approximately 64 cm from the eye-tracker, per Tobii specifications. Each task block began with a 9-point calibration protocol to ensure that the eye-tracker accurately identified the subject's eyes and location of their gaze. A fixation cross was displayed in the middle of the screen for 1 s between trials. During trials, numbers were displayed in black sans-serif text on a grey background and were 2.7 cm in height and 1.9 cm in width if the number was composed of one digit, or 3.8 cm for two digits. There was a 2.7 cm vertical distance between the

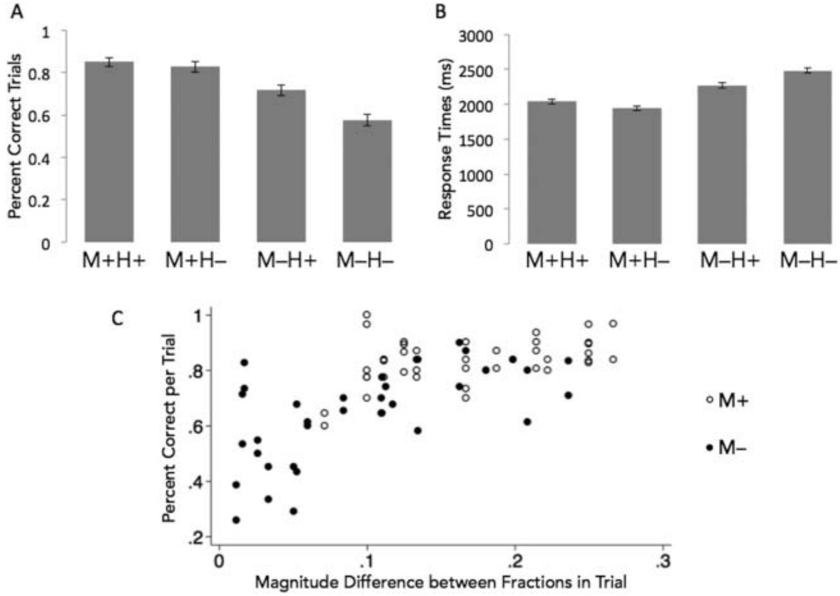


Figure 2. Accuracy (A) and RT on correct trials (B) for each of the four conditions, averaged across the full sample. (C) Average accuracy for each trial displayed as a function of the difference between the magnitudes of the fractions given in the trial for Multiple+ (open circles) and M− (closed circles) conditions.

numbers and the fraction bar, and a horizontal distance of 14.6 cm between the two fractions. Each digit subtended a 2.56-degree visual angle, vertically, with 5.68 degrees between numerator and denominator, and 13.02 degrees between fractions. We spaced the stimuli in such a way as to ensure that viewing each digit required moving one's eyes around the stimulus array, given that the fovea is approximately 2 degrees (Holmqvist et al., 2011).

Data preparation

We designated four primary areas of interest (AOIs) on the fraction comparison task screen, and an additional five AOIs covering the areas between the numbers (Figure 1, panel B). If data from both eyes were available, they were averaged to determine the location of eye fixation; however, a recording from one eye is sufficient to determine the coordinates of a fixation. All contiguous samples whose *x* and *y* coordinates fell within one of the AOIs were collapsed into a single recorded fixation duration. Typical eye fixations last from 100 to 500 ms (Holmqvist et al., 2011). Any fixation that lasted less than 40 ms was presumed to be a transition between AOIs rather than a true fixation, and thus was excluded from further analysis. Any contiguous fixations in the same AOI that were separated by fewer than 100 ms of missing samples were concatenated into a single gaze, as the disruption was presumed to have been caused by a blink.

Because our research questions pertained to differential patterns of eye movements, our primary metric was the number of inter-gaze saccades from one AOI directly to another within each trial. Using the list of gazes generated as described above, we counted the number of saccades between gazes to each of the four primary AOIs. Any two gazes that were separated by more than 300 ms, the duration of a typical blink (Holmqvist et al., 2011), were not counted as a saccade, because it is possible the eyes moved to an intermediate location in that timeframe instead of making a direct saccade. Eye movements from the left to right numerator or vice versa, as well as between denominators, were coded as horizontal saccades. Eye movements between a numerator and denominator in the same fraction were coded as vertical saccades. All additional inter-gaze saccades, such as from one numerator to the opposite fraction's denominator, or those involving non-digit AOIs, were counted as part of the All Saccades metric.

The average All Saccades per trial was 6.43 (SD = 3.06, range 0–30), with an average of 2.54 (SD = 1.72, range 0–11) inter-gaze saccades that were in a broadly horizontal direction, 2.55 (SD = 1.97, range 0–15) saccades that were broadly vertical, and 1.34 (SD = 1.26, range 0–10) saccades that were broadly diagonal. Of those, we chose to analyse only the inter-gaze saccades that moved directly from one AOI to another. We omitted all those that originated or terminated in the interstitial space between them, as we considered this pattern to be a less reliable indicator of numerical comparison. The selected saccades averaged 1.15 horizontal (SD = 1.27, range 0–9) and 1.14 vertical saccades (SD = 1.43, range 0–9) per trial. All trials were included in the analysis of the conditions' effect on behaviour, but incorrect trials (26.5% of the total 2344) as well as trials with no recorded horizontal or vertical saccades (additional 18.9%) were excluded for gaze analyses, leaving 1279 trials. Average All Saccades for included trials was 6.33 (SD = 2.57), with an average 1.43 horizontal and 1.33 vertical inter-gaze saccades ($SD_{horizontal}$ = 1.16; $SD_{vertical}$ = 1.32).

These inter-gaze saccade counts were calculated at the trial level and all analyses were nested within subject, so as to investigate when the participants made more or fewer saccades of a certain type, relative to their general behaviour. We considered collapsing the gaze metrics into proportions of horizontal and vertical saccades per trial, but found those proportions to be misleading when interpreting the analyses. A simple proportion of horizontal-to-vertical saccades led to unusable values when there were no saccades of a particular type in a trial, and we believed those trials to hold valuable information. Using proportions of horizontal (and vertical) to All Saccades led to greater proportions when there were fewer All Saccades in a trial, and lower proportions when there were more All Saccades, even if the number of horizontal saccades was constant across trials. Thus, because All Saccades was an influential metric in its own right, we opted to include it as a controlled variable in regression analyses.

Results

Task performance

Initial analyses confirmed that the task factors impacted both accuracy and RT on correct trials (Figure 2, panels A and B), with participants answering most accurately on M+H+ trials and least accurately on M−H− trials. They responded most quickly on M+H− trials, and slowest on M−H− trials. However, these averages do not reflect how the task conditions affect individual participants' ranges of performance, so all hypotheses were evaluated using mixed linear or logistic regression models at the trial level, in order to test for deviations from participants' average behaviour and performance.

Based on prior research indicating that magnitude difference impacts difficulty of fraction comparisons (Sprute & Temple, 2011), we had ensured that all fraction pairs had a magnitude difference less than 0.3. However, we found that magnitude difference influenced both accuracy and RTs on correct trials even within this restricted range (0.01–0.27). Smaller differences in magnitude between the fractions in a pair had a negative impact on accuracy, with magnitude differences less than 0.1 particularly affected (Figure 2, panel C). A regression of average item accuracy on magnitude difference revealed a positive effect of magnitude difference, $t(5,72) = 3.12$, $p = 0.003$, as well as a negative quadratic effect, $t(5,72) = -2.15$, $p = 0.035$. Despite the fact that the M− trials generally had smaller magnitude differences than the M+ trials, this regression still exhibited an effect of Multiple, $t(5,72) = -2.04$, $p = 0.045$, indicating that the effects of Multiple and magnitude difference were not confounded. Therefore, to control for the effect of magnitude difference while evaluating behavioural effects of the task conditions, as well as the effect of participant skill, we used regression models at the trial level that included fixed Multiple and Half factors, a continuous factor of Magnitude Difference, and a random nested factor by participant.

We first tested our predictions that participants would answer more accurately on M+ and H+ trials than M− and H− trials. A mixed logistic regression yielded a predicted accuracy of 81.5% for M+H+ trials, calculated at the mean value of Magnitude Difference (Table 1). Multiple exerted a main effect: M−H+ trials were associated with 10% lower accuracy, estimated at 71.3%. There was no effect of Half; predicted accuracy on M+H− trials (83.2%) was not different from M+H+ trials. The interaction of the two factors was also not significant: predicted accuracy of M−H− trials was 66.6%. Increasing the distance between the fractions' magnitudes by 0.01 was associated with a 6.2%

Table 1. Multilevel regressions predicting accuracy and response times by trial.

	Accuracy			Response time		
	B (logits)	SE	*Z*	*B*	SE	*Z*
M+H+ (reference)	1.48	0.16	9.43***	2153	76	28.39***
Multiple	−0.57	0.16	−3.59***	195	44	4.44***
Half	0.12	0.17	0.67	−186	43	−4.35***
Interaction: Mult × Half	−0.34	0.21	−1.61	243	63	3.87***
Magnitude difference	0.06	0.01	5.66***	−16	3	−5.46***
Participant (ψ)	0.48	0.09		374		

ψ represents the random effect of individual participants' proficiency, independent of experimental condition.
***$p < 0.001$.

increase in the predicted accuracy. Thus, the two primary influences on accuracy were the Multiple factor and the magnitude difference between the fractions. These data indicate that certain pairs of fractions posed a challenge for the young adults in our study.

Next, we tested how the task factors impacted response times (RTs) on correctly performed trials. The mixed regression model yielded a predicted RT of 2153 ms for M+H+ trials, calculated at the mean value of Magnitude Difference (Table 1). Both Multiple and Half factors exhibited main effects, although in opposite directions: M− trials were slower than the reference group while H− trials were faster. The interaction of M− and H− was also significantly slower than M+H+ trials. Increasing the magnitude difference was associated with faster responses. Thus, RTs were sensitive to both task factors in addition to the effect of Magnitude Difference. However, the effect of Half was in the opposite direction to what we had predicted: we had expected trials in which the fractions were on the same side of 1/2 to be more difficult, as that impeded the use of the benchmarking to 1/2 strategy. Instead, the participants responded more quickly when the fractions were on the same side of 1/2. Both M− trials and smaller magnitude differences were associated with slower RTs. Taken together, these results indicate that all three manipulations influenced how quickly participants could compare fraction magnitudes.

Gaze analyses

To investigate whether the task conditions impacted All Saccades per trial, we used the same type of mixed regression model as for the behavioural measures. Compared to M+H+ trials, there were more saccades on trials that did not have multiples in the denominator (M−H+), and even more on trials without multiples *and* with fractions on the same side of ½ (M−H−). However, there were fewer saccades on M+H− trials, again compared to M+H+ trials (Figure 3, panel A). Additionally, across all trials, there were more saccades when the differences between magnitudes were smaller (see Online Supplementary Table S2 for full regression results). Based on these preliminary

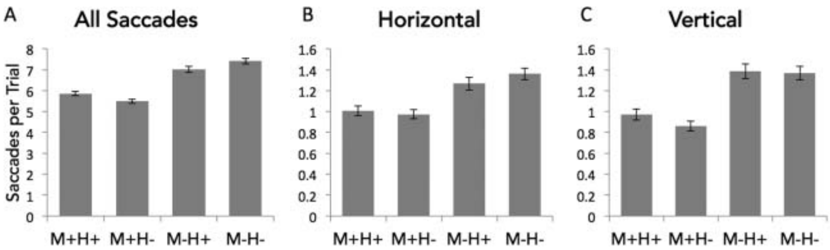


Figure 3. Average number of all saccades (A), horizontal saccades (B), and vertical saccades (C) for each of the four conditions in correct trials containing saccades of interest.

results, we included the All Saccades metric as a regression factor in subsequent gaze analyses, so that our results would indicate where horizontal or vertical saccades were differentially prevalent, and not simply a reflection of more saccades in general. As noted in the "Methods" section, we had considered using a proportion of horizontal-to-vertical saccades, but such an analysis was not optimal given the low number of saccades in this dataset.

We next investigated whether participants adapted their gaze behaviour based on trial type, as could be predicted by a hybrid theory of fraction representation in this task context. We tested for an effect of condition on the number of horizontal and vertical saccades per correct trial (Figure 3, panels B and C). We predicted that M+ trials would be associated with more horizontal saccades than M−, because horizontal saccades indicate componential processing, which is beneficial when one denominator is a multiple of the other (M+). We predicted that H+ trials would be associated with more vertical saccades than H−, because holistic processing is beneficial when magnitude estimates can be benchmarked (H+). These predictions are based on the expectation that people would notice the trial differences and use the most appropriate processing model. Importantly, horizontal and vertical saccades were not equal over all correct trials, with more horizontal on average, $t(1278) = 2.36$, $p = 0.02$. A mixed regression analysis with number of horizontal saccades on correct trials as the outcome measure, the task conditions as factors, and All Saccades per trial as a controlled variable, yielded null results with respect to the task conditions (Table 2). Only All Saccades predicted additional horizontal saccades. This contradicted our hypothesis that participants would make more horizontal saccades on M+ trials. Instead, this finding indicates that participants engaged in componential processing to the same degree whether or not the fraction pairs were conducive to componential processing.

Rather, Multiple impacted the number of *vertical* saccades, with more vertical saccades on M− than M+ trials – the inverse of our hypothesis that there would be more horizontal saccades on M+ than M− trials. Half and

Table 2. Multilevel regressions predicting horizontal and vertical inter-gaze saccades.

	Horizontal saccades			Vertical saccades		
	B	SE	*Z*	*B*	SE	*Z*
M+H+ (reference)	0.36	0.14	2.48*	−0.08	0.16	−0.47
Multiple	0.06	0.07	0.87	0.17	0.08	2.18*
Half	−0.006	0.07	−0.08	−0.03	0.08	−0.36
Interaction: Mult × Half	0.03	0.10	0.27	−0.10	0.11	−0.88
Magnitude difference	0.0004	0.005	0.09	−0.008	0.005	−1.42
All saccades	0.15	0.01	13.31***	0.22	0.009	23.15***
Participant (ψ)	0.60			0.72		

*$p < 0.05$; ***$p < 0.001$.

Magnitude Difference did not exhibit any influence on vertical saccades (Table 2). This result also contradicted our hypothesis that participants would make more vertical saccades on H+ than H− trials. Instead, it suggests that participants engaged in holistic processing when componential processing was not beneficial, but not necessarily in accordance with task conditions that we reasoned should promote holistic processing.

We next sought to test whether the use of componential or holistic processing, as indicated by the number of horizontal or vertical saccades, was associated with consistently accurate or efficient performance, either in general or after taking into account the expected effects of task condition. Whether or not we controlled for difficulty associated with task condition, we found that accuracy did not vary as a function of number of horizontal or vertical saccades (see Table 3). Only Magnitude Difference and All Saccades were associated with accuracy when including all available predictors.

RT for correct trials showed greater sensitivity than did accuracy. In a mixed linear regression model, vertical saccades were associated with slower correct responses, even after controlling for All Saccades and the known effects of task condition (Table 3). Here again, additional vertical saccades were associated not with a strategic benefit, as we expected, but with a performance cost. As an exploratory analysis, we further tested for these effects separately within M+ and M− trials, and found that additional vertical saccades primarily

Table 3. Multilevel regressions predicting accuracy on all trials containing saccades of interest, and response times for correct trials containing saccades of interest.

	Accuracy			Response time		
	B (logits)	SE	*Z*	*B*	SE	*Z*
M+H+ (reference)	2.61	0.22	11.76***	1015	60	17.05***
Multiple	−0.33	0.18	−1.83	78	39	2.05*
Half	0.02	0.20	0.12	−40	38	−1.06
Interaction: Mult × Half	−0.27	0.24	−1.13	127	54	2.38*
Magnitude difference	0.06	0.01	4.53***	−5	3	−1.83
All saccades	−0.19	0.02	−8.01***	179	7	26.91**
Horizontal saccades	0.05	0.05	0.94	21	14	1.48
Vertical saccades	−0.02	0.05	−0.49	33	13	2.53*
Participant (ψ)	0.37	0.09		205		

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

slowed RTs on M+ trials, for which holistic processing was not beneficial (Online Supplementary Table S3).

Our final question was whether participants' performance on this task was associated with their score on our measure of non-numerical reasoning, Analysis-Synthesis (AS) from the Woodcock–Johnson battery. There was no relationship between overall fraction task and AS performance (Pearson's correlation of fraction accuracy with AS: $r = 0.08$, $p = 0.66$; fraction RT with AS: $r = 0.27$, $p = 0.15$). However, when testing the fraction conditions separately, RT for correct trials on the most difficult condition (M−H−) was positively correlated with AS accuracy, $r = 0.40$, $p = 0.03$ (AS is untimed). Thus, participants with better performance on the non-numerical reasoning task exhibited longer RTs on the most difficult condition; this effect was not explained by better reasoners processing the relations more, as indexed by All Saccades per trial. We further tested our theory that componential processing in a mathematical task depends on relational reasoning, but there were no correlations between AS accuracy and participants' average proportion of horizontal saccades per correct trial, either overall ($r = 0.04$, $p = 0.82$) or by condition (all r values < 0.2; all $p > 0.5$).

Discussion

The hybrid theory of the mental representation of fractions posits that people can use either componential or holistic processing when indicated by the task at hand. Prior empirical evidence supports this theory in special cases of the fraction comparison task. In this investigation, we tested the extent to which people's gaze behaviour indicated componential or holistic processing during a task with conditions that subtly privileged one type of processing or the other.

In the condition that promoted componential processing (M+), one denominator was a multiple of the other, in alignment with the commonly taught strategy of converting to equivalent fractions. This strategy reflects componential processing because the primary focus is on the relationship between the fractions' components. M− trials had denominators that were not multiples of each other, making componential processing more difficult. In the condition that privileged holistic processing (H+), the fractions' magnitudes were on opposite sides of $\frac{1}{2}$, in alignment with the commonly used strategy of benchmarking. This strategy reflects holistic processing because the primary focus is on the mathematical relationship of numerator to denominator within fractions. In H− trials, both fractions were either greater than or less than $\frac{1}{2}$, making holistic processing more difficult.

We assessed how participants approached the task by calculating the number of horizontal eye movements, an indicator of componential processing, and the number of vertical eye movements, an indicator of holistic processing, on each correct trial. We then tested our hypotheses that participants would adjust their approach according to task condition, and that condition-aligned eye movements would be associated with better and faster performance.

We found evidence that people did in fact adjust their processing approaches in response to trial type, but not in the manner we hypothesised. In accordance with hybrid theory, we expected to see evidence of both horizontal and vertical saccades. We did, but the balance was tipped towards horizontal saccades, indicating a greater prevalence of componential processing overall. We had predicted that participants would demonstrate more componential processing on M+ than M− trials, and more holistic on H+ than H− trials. Instead, participants showed the same degree of componential processing regardless of trial type, and more holistic processing on M− than M+ trials.

Taken together, these results are compatible with the interpretation that instead of adjusting their approach to benefit from the affordances of the trial, participants generally attempted to use componential processing, and adjusted their approach to use holistic processing when componential processing was not beneficial. Performance data supported this interpretation: participants performed much better on M+ than M− trials, and there was no accuracy difference between H+ and H− trials. This conclusion has been previously proposed in studies including shared components (Meert et al., 2009; Obersteiner & Tumpek, 2016), but we did not expect it in a solely mixed-pairs context because the shared components trials that promoted componential processing were not included here.

Indeed, prior studies with mixed-pair stimuli led to predominantly holistic processing (Ischebeck et al., 2016; Obersteiner & Tumpek, 2016), while we found componential processing to be more generally prominent. This difference from preceding findings also contradicts the prediction suggested by relational complexity theory, which states that computing the magnitude of a fraction reduces its complexity from a relation to a single entity, and thereby makes the subsequent judgement between two fractions a simple comparison rather than a second-order relation. Based on that theory, holistic processing should always be more efficient than componential when comparing mixed-pair fractions, yet our evidence suggests that people engaged in componential processing to a greater extent than holistic.

One possible explanation for this unexpected finding is that people simply did not notice the opportunity to use a benchmarking strategy when it was available, but did notice that one denominator was sometimes a multiple of the other, and so consistently applied strategies that align with componential processing. The fact that there was no difference in accuracy between the H+ and H− trials supports this interpretation; had they used a benchmarking to $\frac{1}{2}$ strategy, they should have performed better on H+ than H− trials.

Furthermore, participants were quicker to respond correctly on H− trials than H+, opposite to our hypothesis that H+ trials would be easier to evaluate. It could be the case that instead of benchmarking to $\frac{1}{2}$, participants estimated which of the two fractions was farther from 0, in the case of smaller fractions, or closer to 1, in the case of larger ones. If so, the H− trials, in which both fractions were either greater or less than $\frac{1}{2}$, would be easier than H+. However, the lack of effect of the Half manipulation on accuracy casts doubt that an alternate benchmarking strategy is the full explanation.

Another possible explanation for the prevalence of componential processing is that the higher order relational comparison required for this task may be less cognitively demanding than the mathematical estimation of magnitude required by holistic processing. In this mixed-pairs context, estimating the fractions' magnitudes is always more efficient from a relational complexity perspective, but that assumes the mathematical procedures for each step are equivalent in difficulty. It may be the case that finding mathematical relations between fractions is easier than estimating magnitudes within fractions. Given that our sample was from a highly selective university, and that most of our participants rated themselves as confident in their mathematics abilities, we expected that they would be proficient in estimating fraction magnitudes. However, the reliance on componential processing suggests that participants are more comfortable with higher order relational reasoning than they are with mathematical calculations. Whether this is because they found the relational comparisons easier to make, or because they were executing an overlearned procedure instead of evaluating the trial affordances, would be interesting to discover. It may be possible to evaluate this question by modifying the experimental paradigm to give unlimited time per trial. We set a conservative time limit of 4 s per trial in order to motivate efficient processing; however, in a recent study involving young adults performing a similar paradigm, average RTs were 4.7 s for mixed-pair trials (Obersteiner & Tumpek, 2016), so our response window was perhaps slightly too stringent. It is possible that placing time pressure on participants led them to stick with a singular approach instead of taking the time to choose the optimal strategy for each trial.

Another important extension of this work would be to conduct a scan path analysis investigating the order of saccades within a trial, to better characterise strategy use (Hayes & Henderson, 2017). Indeed, different strategies would be expected to be associated with different sequences of eye movements; for example, an equivalent fractions strategy might be characterised by more horizontal saccades at the beginning of a trial than the end. We did not have sufficient saccades per trial in the current dataset to conduct this analysis; furthermore, Ischebeck and colleagues found initial scan patterns to be somewhat idiosyncratic (Ischebeck et al., 2016), so it seems that a much larger sample and modified paradigm would be needed to further examine strategy use.

To facilitate scan path analysis, future studies should make several modifications to this mixed-pair fraction comparison paradigm. One modification would be to ensure that peripheral vision is not sufficient to encode the stimuli. We designed the screen to maintain familiar fraction notation, but we believe that this design, as well as our conservative accounting of saccades, resulted in a relatively low per-trial saccade count. We also defined narrow AOIs, counted only saccades in which the eyes moved directly from one number to another, and excluded any movements that included a fixation on the fraction bars or central space, or a latency longer than 300 ms. We took these measures to boost confidence in our findings, and to focus our analyses on specific gaze patterns, but it is possible that less conservative data selection would lead to additional insights.

In addition to investigating gaze behaviour entirely within a mixed-pairs context, this study extended prior methods by analysing the effect of gaze behaviour on performance above and beyond the effect of task condition. Regarding our hypothesis that condition-aligned gaze behaviour would be associated with better and faster performance, we predicted that M+ trials with more horizontal saccades were more likely to be correct with a faster RT, and similar for H+ trials with more vertical saccades. Instead, we found the inverse: that gaze behaviour not aligned to condition was associated with poorer performance, even when controlling for task difficulty. Specifically, a greater number of vertical saccades on M+ trials were associated with slower correct responses. Therefore, when participants used holistic processing on M + trials, which were designed to promote componential processing, it did not affect accuracy but did slow response times.

The fact that gaze behaviour provides additional explanation of performance variation on trials, beyond the known variance due to task condition, underscores the utility of gaze metrics as a supplement to behavioural data. Here, the gaze data provides insight into participants' mental processes as they struggle with a problem. Additionally, although there was a range of accuracy between participants, our analyses tested for deviations from an individual's average performance, and thus the RT metric is more sensitive; it may be the case that accuracy differences are detectable in a broader sample of participants. Future research with different populations and more granular analyses may provide even more insights into problem-solvers' attention to the different types of relational information present in the task, and how that attention serves or detracts from their performance.

Our expectation that participants would attend to multiple sets of relations underpinned our hypothesis that their performance on this task would be associated with their performance on a test of non-numerical reasoning. However, this was not the case. Performance on the non-numerical test of reasoning was associated only with response times on the most difficult condition of this fraction comparison task, and not with our gaze-based measures of processing types. The non-numerical reasoning task required participants to carefully examine both the problem and the given reference key. Similarly, the most difficult condition of the fraction task obstructed both of the processing approaches participants may use and thus required more time and effort than the other conditions. Although this result was unexpected, it is compatible with the idea that participants who were more accurate at non-numerical reasoning also attended more closely to numerical relationships in mathematical reasoning contexts. Judging from overall number of saccades, it was not the case that these participants simply made more relational comparisons. However, the correlation between these tasks was not explained by our planned gaze metrics, and so we cannot draw a conclusion about whether a particular type of mathematical reasoning is related to non-numerical reasoning. Future studies including additional gaze metrics, as well as additional cognitive measures, should further elucidate whether or how mathematical reasoning relies on more general relational thinking (English & Halford, 1995; Goswami, 2004).

Notably, this study was conducted with high-performing young adults. It may well be that we were only able to detect a relationship between fraction comparison and our measure of relational reasoning on the most difficult condition because it displayed sufficient variability in performance. We predict a stronger relationship among participants who vary more widely in mathematical ability.

A limitation of this study that would be especially important to address in a wider sample is that we did not collect survey data about which mathematical strategies the participants recall learning from their elementary curriculum. It is possible our participants used strategies that we did not assess. Future studies should also expand this research to a wider range of participants. In this initial test of the mixed-pairs task, we sampled from a highly selective university. None of the participants performed at ceiling, and there was a sufficient range of variability to test our hypotheses. However, including lower performing adults and children is an important next step in understanding the link between mathematical and relational reasoning and how it is successfully or unsuccessfully developed.

In closing, this study provides evidence that people engage in both componential and holistic processing, even in a context in which none of the fraction pairs shared components. Our findings also supported the idea that gaze data can supplement behavioural analyses by revealing participants' approaches when performance is sub-optimal. Furthermore, our findings suggest that our participants favoured higher order relational reasoning over mathematical calculations, which may be an initial indication of how relational reasoning supports mathematical skill, but also indicates that a more flexible conceptualisation of fractions, as both magnitudes and relations between components, would enable them to choose the optimal approach in different cases. Further investigations with a modified version of this paradigm may better characterise successful strategies and further elucidate how relational thinking supports mathematical cognition.

Acknowledgments

We thank the research assistants who contributed to collecting these data: Maia Barrow, Heather Anderson, Jesse Niebaum, Jeffrey Crawford, Maddy Hubbard, Ryunosuke Fujinomake, Atiya Dphrepaulezz, Vinicius Marinho, Jennifer Bob, Hannah Bystritsky, Sara Isakovic, Andrew Shu, Shree Patel, and Estefania Pulido.

Disclosure statement

The authors declare no conflicts of interest.

Funding

This research was supported by a James S. McDonnell Foundation Scholar Award, and a Jacobs Foundation Scholar Award to Silvia A. Bunge; and the Institute of Education Sciences predoctoral training grant [grant number R305B090026] to the University of California, Berkeley.

References

Bonato, M., Fabbri, S., Umilta, C., & Zorzi, M. (2007). The mental representation of numerical fractions: Real or integer ? *Journal of Experimental Psychology. Human Perception and Performance, 33*(6), 1410–1419. doi:10.1037/0096-1523.33.6.1410

Clarke, D. M., & Roche, A. (2009). Students' fraction comparison strategies as a window into robust understanding and possible pointers for instruction. *Educational Studies in Mathematics, 72*(1), 127–138. doi:10.1007/s10649-009-9198-9

English, L. D., & Halford, G. S. (1995). *Mathematics education: Models and processes*. Mahwah, NJ: Lawrence Erlbaum Associates.

Faulkenberry, T. J., Montgomery, S. A., & Tennes, S.-A. N. (2015). Response trajectories reveal the temporal dynamics of fraction representations. *Acta Psychologica, 159*, 100–107. doi:10.1016/j.actpsy.2015.05.013

Faulkenberry, T. J., & Pierce, B. H. (2011). Mental representations in fraction comparison: Holistic versus component-based strategies. *Experimental Psychology, 58*(6), 480–489. doi:10.1027/1618-3169/a000116

Fazio, L. K., DeWolf, M., & Siegler, R. S. (2016). Strategy use and strategy choice in fraction magnitude comparison. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 42*(1), 1–16.

Gabriel, F. C., Szucs, D., & Content, A. (2013). The development of the mental representations of the magnitude of fractions. *PloS One, 8*(11), e80016. doi:10.1371/journal.pone.0080016

Goswami, U. (2004). Commentary: Analogical reasoning and mathematical development. In L. D. English (Ed.), *Mathematical and analogical reasoning of young learners* 169–186. Mahwah, NJ: Lawrence Erlbaum Associates.

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

Halford, G. S., & Wilson, W. H. (1980). A category theory approach to cognitive development. *Cognitive Psychology, 12*, 356–411.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *The Behavioral and Brain Sciences, 21*(6), 803–864. doi:10.1017/S0140525X98001769

Hayes, T. R., & Henderson, J. M. (2017). Scan patterns during real-world scene viewing predict individual differences in cognitive capacity. *Journal of Vision, 17*(5), 1–17. doi:10.1167/17.5.23.doi

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.

Ischebeck, A., Schocke, M., & Delazer, M. (2009). The processing and representation of fractions within the brain: An fMRI investigation. *NeuroImage, 47*, 403–413. doi:10.1016/j.neuroimage.2009.03.041

Ischebeck, A., Weilharter, M., & Körner, C. (2016). Eye movements reflect and shape strategies in fraction comparison. *Quarterly Journal of Experimental Psychology, 69*(4), 713–727. doi:10.1080/17470218.2015.1046464

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8*, 441–480. doi:10.1016/0010-0285(76)90015-3

Kafkas, A., & Montaldi, D. (2012). Familiarity and recollection produce distinct eye movement, pupil and medial temporal lobe responses when memory strength is matched. *Neuropsychologia, 50*(13), 3080–3093. doi:10.1016/j.neuropsychologia.2012.08.001

McGrew, K. S., & Hessler, G. L. (1995). The relationship between the WJ-R Gf-Gc cognitive clusters and mathematics achievement across the life-span. *Journal of Psychoeducational Assessment, 13*(1), 21–38. doi:10.1177/073428299501300102

Meert, G., Grégoire, J., & Noël, M.-P. (2009). Rational numbers: Componential versus holistic representation of fractions in a magnitude comparison task. *Quarterly Journal of Experimental Psychology*, *62*(8), 1598–1616. doi:10.1080/17470210802511162

Meert, G., Grégoire, J., & Noël, M. P. (2010). Comparing 5/7 and 2/9: Adults can do it by accessing the magnitude of the whole fractions. *Acta Psychologica, 135*(3), 284–292. doi:10.1016/j.actpsy.2010.07.014

Miller Singley, T. A., & Bunge, S. A. (2014). Neurodevelopment of relational reasoning: Implications for mathematical pedagogy. *Trends in Neuroscience and Education, 3*(2), 33–37. doi:10.1016/j.tine.2014.03.001

Morsanyi, K., Devine, A., Nobes, A., & Szűcs, D. (2013). The link between logic, mathematics and imagination: Evidence from children with developmental dyscalculia and mathematically gifted children. *Developmental Science, 16*(4), 542–553. doi:10.1111/desc.12048

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature, 215*(5109), 1519–1520. doi:10.1038/2151519a0

Ni, Y., & Zhou, Y.-D. (2005). Teaching and learning fraction and rational numbers: The origins and implications of whole number bias. *Educational Psychologist, 40*(1), 27–52. doi:10.1207/s15326985ep4001

Obersteiner, A., & Tumpek, C. (2016). Measuring fraction comparison strategies with eye−tracking. *ZDM Mathematics Education, 48*, 255–266. doi:10.1007/s11858-015-0742-z

Prado, J., Van der Henst, J.-B., & Noveck, I. A. (2008). Spatial associations in relational reasoning: Evidence for a SNARC-like effect. *Quarterly Journal of Experimental Psychology*, *61*(8), 1143–1150. doi:10.1080/17470210801954777

Primi, R., Ferrão, M. E., & Almeida, L. S. (2010). Fluid intelligence as a predictor of learning: A longitudinal multilevel approach applied to math. *Learning and Individual Differences, 20*(5), 446–451. doi:10.1016/j.lindif.2010.05.001

Richland, L. E., & Begolli, K. N. (2016). Analogy and higher-order thinking: Learning mathematics as an example. *Policy Insights from the Behavioral and Brain Sciences, 3*(2), 160-168. doi:10.1177/2372732216629795

Shepherd, M., Findlay, J. M., & Hockey, R. J. (1986). The relationship between eye movements and spatial attention. *The Quarterly Journal of Experimental Psychology, 38*(3), 475–491. doi:10.1080/14640748608401609

Smith, J. P., III (1995). Competent reasoning with rational numbers. *Cognition and Instruction, 13*(1), 3–50.

Sprute, L., & Temple, E. (2011). Representations of fractions: Evidence for accessing the whole magnitude in adults. *Mind, Brain, and Education, 5*(1), 42–47. doi:10.1111/j.1751-228X.2011.01109.x

Vendetti, M. S., Starr, A., Johnson, E. L., Modavi, K., & Bunge, S. A. (2017). Eye movements reveal optimal strategies for analogical reasoning. *Frontiers in Psychology, 8*, 1–9. doi:10.3389/fpsyg.2017.00932

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III tests of cognitive abilities*. Itasca, IL: Riverside Publishing.

Zamarian, L., Ischebeck, A., & Delazer, M. (2009). Neuroscience of learning arithmetic–evidence from brain imaging studies. *Neuroscience and Biobehavioral Reviews, 33*(6), 909–925. doi:10.1016/j.neubiorev.2009.03.005