

## Empirical Research

# Eye Gaze Patterns Reflect How Young Fraction Learners Approach Numerical Comparisons

Alison T. Miller Singley<sup>\*a</sup>, Jeffrey Lynn Crawford<sup>ab</sup>, Silvia A. Bunge<sup>ac</sup>

[a] Department of Psychology, University of California at Berkeley, Berkeley, CA, USA. [b] Department of Psychological Sciences, University of Connecticut, Storrs, CT, USA. [c] Helen Wills Neuroscience Institute, University of California at Berkeley, Berkeley, CA, USA.

## Abstract

Learning fractions is notoriously difficult, yet critically important to mathematical and general academic achievement. Eye-tracking studies are beginning to characterize the strategies that adults use when comparing fractions, but we know relatively little about the strategies used by children. We used eye-tracking to analyze how novice children and mathematically-proficient adults approached a well-studied fraction comparison paradigm. Specifically, eye-tracking can provide insights into the nature of differences: whether they are quantitative—reflecting differences in efficiency—or qualitative—reflecting a fundamentally different approach. We found that children who had acquired the basic fraction rules made more eye movements than did either adults or less proficient children, suggesting a thorough but inefficient problem solving approach. Additionally, correct responses were associated with normative gaze patterns, regardless of age or proficiency levels. However, children paid more attention to irrelevant numerical relationships on conditions that were conceptually difficult. An exploratory analysis points to the possibility that children on the verge of making a conceptual leap attend to the relevant relationships even when they respond incorrectly. These findings indicate the potential of eye-tracking methodology to better characterize the behavior associated with different levels of fraction proficiency, as well as to provide insights for educators regarding how to best support novices at different levels of conceptual development.

**Keywords:** eye-tracking, math, fractions, learning, cognitive development

Journal of Numerical Cognition, 2020, Vol. 6(1), 83–107, <https://doi.org/10.5964/jnc.v6i1.119>

Received: 2017-02-01. Accepted: 2019-04-29. Published (VoR): 2020-06-15.

\*Corresponding author at: Department of Psychology, University of California, Berkeley, 2121 Berkeley Way, Room 3302, Berkeley, CA 94720-1650, USA. E-mail: [atms@berkeley.edu](mailto:atms@berkeley.edu)



This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License, CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The ability to understand and interpret fractions is an important foundational skill for the development of mathematic ability. Just as algebra competency has been shown to be the gateway into careers in math and science (National Mathematics Advisory Panel, 2008), fraction competency seems to be the gateway into understanding algebra (Booth & Newton, 2012). However, algebra teachers in the United States rate their students as having extremely poor knowledge of rational numbers, nearly the weakest topic of 15 core mathematical areas (Hoffer, Venkataraman, Hedberg, & Shagle, 2007). With deficits in fraction competency seeming endemic, identifying possible pedagogical improvements is an imperative.

There are many conceptual challenges associated with the learning of fractions. One of the biggest hurdles is that the magnitude of a fraction is *not* defined by the values of the component numbers, but instead by the relationship between the numerator and denominator. Consequently, fractions must be mentally represented

either as single integrated magnitudes, called holistic representation, or as their component numbers, referred to as componential representation. Each representation is valuable in certain contexts, and prior research has found evidence supporting both mental representations (e.g., Bonato, Fabbri, Umiltà, & Zorzi, 2007; Ischebeck, Schocke, & Delazer, 2009; Meert, Grégoire, & Noël, 2009; Obersteiner et al., 2014).

Studies using the fraction comparison paradigm, in which participants are asked to select the fraction with the larger magnitude, have found that adults adaptively use both componential and holistic representations, depending on the task context (Huber, Moeller, & Nuerk, 2014; Meert et al., 2009). For example, if the fractions presented share the same denominator (e.g.,  $3/5$  and  $4/5$ ), a simple heuristic of selecting the larger numerator will yield the correct answer, and behavioral evidence aligns to the theory of componential representation (Ischebeck et al., 2009; Meert et al., 2009). However, when the pair shares no common components (e.g.,  $3/5$  and  $4/6$ ), it may be necessary to calculate the integrated magnitude or use an alternate strategy in order to select the larger fraction. Accordingly, in this context, evidence aligns with a holistic magnitude representation of fractions (Meert, Grégoire, & Noël, 2010a; Obersteiner et al., 2014; Sprute & Temple, 2011). When the task interleaves both shared-components and mixed pair trials, adults select an appropriate strategy after a brief scan of the problem space, and focus on the relevant components of the comparison (Ischebeck et al., 2016). There is even evidence for the influence of both component and magnitude representations within the same trial, with components being accessed automatically and magnitudes after processing (Faulkenberry, Montgomery, & Tennes, 2015). Furthermore, better performance on the fraction comparison task is associated with use of a wider range of strategies (Fazio, DeWolf, & Siegler, 2016).

## Fraction Comparison by Young Learners

The above-mentioned findings centered on adults who demonstrated a certain degree of proficiency. Fraction comparison studies with child participants are much fewer, and have shown, as with adults, that the more successful participants use a wider range of strategies, and select strategies that align to the particular task challenges (Clarke & Roche, 2009; Smith III, 1995). Yet even these studies focused on children who had completed at least several years of fraction education. In the current study we selected participants who had some familiarity with unit fractions (i.e.,  $1/n$ ) but had not yet had much formal education, as we sought to investigate how novices approach these complex mathematical problems.

The ability to solve complex problems, in any context, by identifying patterns across multiple sets of mental representations is called relational reasoning (Dumas, Alexander, & Grossnickle, 2013). Examples of relational reasoning include the ability to solve propositional analogies (e.g., puppy is to dog as kitten is to cat), as well as matrix reasoning and transitive inference. A wide-ranging body of research has shown that domain-general executive skills, and particularly relational reasoning, support academic success. Relational reasoning has long been cited as a predictor of future academic achievement (e.g., Gottfredson, 1997) and has more recently been linked specifically to the acquisition of mathematical skills (Green, Bunge, Briones Chiongbian, Barrow, & Ferrer, 2017; Primi, Ferrão, & Almeida, 2010). Although the mechanism of learning is not well characterized, several researchers posit that relational reasoning supports mathematical thinking through attention to structural similarities between familiar and novel problems (Miller Singley & Bunge, 2014; Richland & Begolli, 2016; Richland & McDonough, 2010).

Attending to the structure of the fraction comparison task highlights the complexity of the relationships between the four numbers in each problem. While the prompt is always the same—*Select the larger of two fractions*—different relationships are relevant in different types of problems. In the simplest case, when the two denominators are equal, the task requires only a single comparison between the two numerators. And this comparison is familiar to young learners: the fraction with the larger numerator is the one with the larger value. Thus, extending the known mathematical rule that higher numbers indicate larger magnitudes is helpful in this simplest case.

The converse case, in which the numerators are equal and the denominators differ, requires comparison between the denominators. However, extending familiar information to this novel problem leads to the incorrect response. Instead of selecting the larger number, the students must learn a new rule: that the smaller denominator indicates the larger fractional value. According to Dumas et al. (2013), antithetical, or oppositional, reasoning is a type of relational reasoning, but it is practiced far less often in formal educational settings. The smaller-denominator rule in particular has been posited to be a transitional step toward comprehensive fraction understanding (Rinne, Ye, & Jordan, 2017), as there is evidence that a smaller denominator causes a “Stroop-like” interference (Meert, Grégoire, & Noël, 2010b). These studies further illustrate how expanding children’s understanding of relational rules may improve their skill with fractions.

In the most complex case of the fraction comparison task, all four numbers are different, and, depending on the task affordances, a variety of strategies may be useful. The relationship between the numerator and denominator of each fraction defines its value, and so attending to the integrated magnitudes and then comparing them will reliably produce the correct answer. However, this strategy is both conceptually and mathematically challenging: it requires proficiency in both calculation and relational reasoning. From a relational reasoning perspective, this strategy has the same problem structure as traditional analogies. It is a second-order comparison, or the comparison of two first-order relationships, which is more cognitively taxing than simple comparisons (Halford, Wilson, & Phillips, 1998). Children are known to spontaneously use analogical thinking in their learning (Inagaki & Hatano, 1987; White, Alexander, & Daugherty, 1998), but analogies are rarely used effectively in formal mathematics education (Richland, Holyoak, & Stigler, 2004).

The capacity for relational reasoning improves through middle childhood (Bazargani, Hillebrandt, Christoff, & Dumontheil, 2014; Halford et al., 1998; Wendelken et al., 2018). Because the representation of 2<sup>nd</sup>-order relations is challenging, many students learn specific strategies to handle the mixed-pair fraction comparisons, such as converting to like denominators (i.e., multiplying one fraction by  $n/n$  such that the denominators become equal and the numerator comparison becomes straightforward), or cross-multiplying (i.e., multiplying each numerator by the opposite denominator and comparing the products, which is a simplified algorithmic method of converting to equivalent denominators). More experienced learners may look at the holistic magnitude when necessary (Obersteiner et al., 2014), or may continue to use these specific strategies when warranted (Faulkenberry & Pierce, 2011). Thus, in the mixed pair case of this task, each pairwise relationship between all four numbers may be useful to consider, but some are more familiar and thereby more accessible to new learners than others.

Just as relational reasoning develops throughout childhood, so do several additional cognitive skills that undergird performance on the fraction comparison task. In particular, the ability to flexibly apply different mathematical rules to different cases, or cognitive flexibility, as well as processing speed, both improve through

adolescence (e.g., Davidson, Amso, Anderson, & Diamond, 2006; Diamond, 2002; Luna, Garver, Urban, Lazar, & Sweeney, 2004). Additionally, working memory span, or the number of pieces of information one can keep in mind simultaneously, improves through mid-childhood (Gathercole, 1999; Perone, Simmering, & Spencer, 2011).

In summary, while adults can recognize the different cases within the fraction comparison task and modify their strategies accordingly, the task is much more difficult for children. Not only are they new to working with fractions, but their relational reasoning, task-switching, and working memory skills are all less efficient than those of adults. Given their status as novice learners, we sought to investigate whether or how children approached the fraction comparison task differently from adults.

## Fraction Comparison and Eye-Tracking

The aforementioned studies have used highly precise behavioral and chronometric methods to make inferences about mature and developing mental representations of fractions, but it is difficult to gain insights about the variety of strategies that people employ without repeatedly asking for verbal reports while they solve problems, which incurs the risk of influencing their approach. However, eye-tracking technology can be used to track people's eyes as they examine a problem. Eye gaze is intimately related to attention (e.g., Deubel & Schneider, 1996; Shepherd, Findlay, & Hockey, 1986), and therefore we can infer a person's strategy by tracking their eye fixations and eye movements, or saccades (Grant & Spivey, 2003). Pairing eye gaze metrics with quantitative metrics that reflect efficiency of cognitive processing allows researchers to distinguish between quantitative group differences—that reflect proficiency with cognitive functions underlying a particular task—and qualitative differences—that reflect fundamentally different strategies or approaches to the task. This distinction was not possible solely with behavioral methods.

Eye-tracking studies using the fraction comparison paradigm leveraged patterns of saccades between the numbers displayed on the screen to infer a person's strategy. Obersteiner and Tumpek (2016) and Ischebeck et al. (2016) both found that when people compared fraction pairs with the same denominator (e.g.,  $3/5$  and  $4/5$ ), saccades between numerators were more prevalent, whereas when comparing fraction pairs with the same numerator, saccades between denominators were more prevalent (e.g.,  $4/5$  and  $4/6$ ). Obersteiner and Tumpek (2016) additionally found that saccades between the numerator and denominator within the same fraction were more common when the fractions shared no common components. These initial eye-tracking findings lend support to the hybrid theory of mental representation of fractions, as they show that adults use componential strategies when they are adaptive, and holistic strategies when all digits need to be taken into account. The fraction comparison studies involving children have used interview techniques to elaborate the various strategies employed (e.g., Clarke & Roche, 2009; Smith III, 1995). To our knowledge, however, none have probed strategic approaches using eye-tracking.

Eye-tracking methodology has illuminated different strategies in use for different task conditions, but also in different groups of people. A set of studies using the number line magnitude placement task documented the use of less and more sophisticated strategies in children (Schneider et al., 2008), adults (Sullivan, Juhasz, Slattery, & Barth, 2011), and atypically developing children (van't Noordende, van Hoogmoed, Schot, & Kroesbergen, 2016). When placing a random number on a 0–100 number line, novices tended to look primarily at the endpoints and midpoint of the line, while participants who were older and more skilled seemed to divide the line

into finer segments and looked preferentially at more precise benchmarks. This set of findings highlights the possibility that the mathematical strategies used by children as they are learning new concepts are qualitatively different than those used by experienced adults.

Beyond these mathematical tasks, eye-tracking research has also identified some general differences in strategic approach due to differing skill levels. A meta-analysis of proficiency studies (Gegenfurtner, Lehtinen, & Säljö, 2011) reported that experts in a variety of professional arenas had shorter fixation durations, more fixations on task-relevant areas, fewer fixations on task-redundant areas, longer saccades, and shorter times to first fixate on relevant information. Our version of the fraction comparison task gives the opportunity to demonstrate many of these behaviors, in that there are task-relevant and task-redundant areas of the screen, and it requires knowledge of specific fraction rules and strategies. Thus, we expected the differences in knowledge between children and adults to be reflected in qualitatively different eye movements.

In addition to providing insights into problem-solving approaches, eye-tracking metrics can also capture quantitative differences related to efficiency of cognitive processing, thereby allowing us to discern whether group differences are qualitative or quantitative. Eye-tracking research has shown that children generally respond more slowly to stimuli than do adults (e.g., Bucci & Seassau, 2012). Working memory tasks elicit pupillary responses, detectable with eye-tracking methodology, that differ between children and adults (Johnson, Miller Singley, Peckham, Johnson, & Bunge, 2014; Luna et al., 2004). These general cognitive skills tend to improve with maturation, so we expected to capture quantitative differences in the number of eye movements between children and adults.

## Current Study

In this study we sought to identify the qualitative and quantitative differences in problem-solving approaches between new learners and mathematically-proficient adults. We compared the performance and gaze behavior of adults to those of fifth graders (9–11 year olds) near the beginning of the school year, on a fraction comparison task that included both mixed pairs and pairs with same components. Both groups completed the identical task while we measured their behavioral performance and tested for differences in their eye movements. We measured both raw numbers of saccades, which reflect cognitive efficiency, and percentages of particular types of saccades per trial, which reflect qualitative patterns of gaze behavior and indicate problem-solving strategy.

Based on the research described above showing a general improvement in cognitive skills with age, we predicted that adults would demonstrate higher efficiency on the fraction comparison task, as evidenced by fewer overall saccades across task conditions. Although children might take longer and exhibit more saccades overall, we predicted that saccade patterns, that is, the relative number of different types of saccades, would be related to mathematical proficiency. Thus, we predicted that children would exhibit qualitatively similar accuracy and gaze patterns to adults on the simpler cases, which they may be familiar with, and poorer performance and disorganized gaze behavior on more complex cases that they have yet to learn.

## Method

### Participants

We recruited 35 5<sup>th</sup>-grade children (ages 9–11) and 38 college students (ages 18–22) for this study. The children were recruited from a charter school in a socioeconomically depressed community in Oakland, California. 95% of students at this school are eligible for free or reduced price lunch. Academically, only 23% meet state literacy goals (compared to 44% in the state overall), and only 25% meet state mathematics goals (compared to 33% in the state). The child participants completed this study as part of an effort to assess the cognitive benefits of chess training. The young adult participants consisted of undergraduate students at the University of California at Berkeley who participated in the study for course credit in a Psychology course, as part of a larger study on adults' fraction strategies. All study procedures were approved by the Committee for the Protection of Human Subjects at the University of California at Berkeley.

Three children were excluded from the study on the basis of less than 50% valid eye gaze data. Two adults and three children were excluded for poor performance based on the clustering procedure described below. The final sample included 29 children ( $M_{\text{Age}} = 10.6$ ,  $SD = 0.55$ ; 17 girls, 12 boys) and 36 young adults ( $M_{\text{Age}} = 20.4$ ,  $SD = 1.2$ ; 24 women, nine men, three declined to state). All participants had normal or corrected-to-normal vision.

### Procedure

Children were given permission to leave class, and were brought to a Tobii eye-tracker that was set up in a quiet room inside the school for a 20-minute eye-tracking session that included this task after completing a working memory task and, last, a resting scan. Adults visited the lab for a 1-hour session that included a different battery of tasks: this task was the first, followed by a more difficult version of the fraction comparison task, a paper-and-pencil test of relational reasoning, a version of fraction comparison that contained proper and improper fractions, and a final strategy interview.

Participants were told that they would see two fractions on the screen, and that they would need to decide as quickly as they could which fraction represented the larger magnitude, entering their choice by pressing the left or right arrow key on a standard computer keyboard. They were not instructed to use any particular strategy in solving the fraction comparison problems, nor were they given any feedback during the trials. The trials commenced immediately without any practice trials. The experiment lasted approximately 5 minutes. Trials were self-paced, with a limit of 8 seconds, and a fixation cross was presented for one second between successive trials.

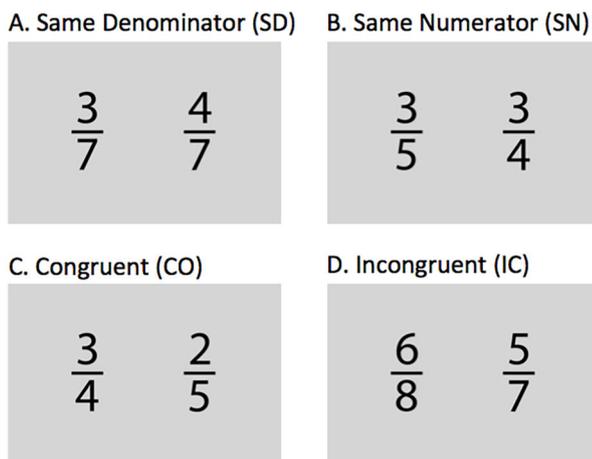
The experiment was conducted on a Tobii T120 eye-tracker, with a sampling rate of 120 Hz (one measurement every 8.3 milliseconds). Participants were asked to sit in front of the eye-tracker at the recommended distance of approximately 64 cm. The session began with a 9-point calibration protocol to ensure that the eye tracker accurately identified the participant's eyes and location of their gaze.

During the task session, two fractions were shown side by side on the screen, each digit subtending 2.2 horizontal degrees  $\times$  3.4 vertical degrees, with a visual angle of 8.51° between fractions and 1.71° between numerators and denominators. The digits in this version were placed with less vertical separation than in other

studies (e.g., [Ischebeck et al., 2016](#)), but because our participants were just learning fractions we wanted to ensure they appeared in a recognizable format. Because the fovea typically extends 2° ([Holmqvist et al., 2011](#)), this layout may have enabled participants to encode the stimuli using peripheral vision rather than having to foveate each one, thereby resulting in fewer saccades between stimuli.

There were 32 trials total, divided into four interleaved conditions with eight fraction pairs each, adapted from [Ischebeck et al. \(2009\)](#). In these eight fraction pairs, four were unique pairs and the other four were reversed duplicates of the first four, to counterbalance the correct responses between left and right.

Following [Ischebeck et al. \(2009\)](#), we used four conditions that elicit distinct behavioral signatures. In the Same Denominator (SD) condition, fraction pairs had the same denominator but different numerators ([Figure 1a](#)). This was the simplest condition, because when two fractions have the same denominator, then the larger fraction has the larger numerator, in alignment with the rules of counting numbers. In the Same Numerator (SN) condition, each of the fraction pairs had different denominators, but the same numerators ([Figure 1b](#)). These fraction pairs are solved by knowing that, if the two numerators are the same, the one with the *smaller* denominator is the larger fraction. The third condition, called the congruent condition (CO), was a direct extension of the SN and SD conditions, meaning a decision based on either numerators or denominators would lead to a correct response: the correct answer had both a larger numerator and a smaller denominator ([Figure 1c](#)). The most difficult condition was the incongruent condition (IC), in which one fraction had both a larger numerator and a larger denominator, providing inconsistent cues, such that all four digits had to be considered to select the correct response ([Figure 1d](#)). Conditions were interspersed pseudo-randomly over the course of a single block of trials.



*Figure 1.* Sample items for each task condition.

The numbers depicted in the fractions were single digits between one and nine, so that the stimuli would be highly familiar to both children and adults (see stimulus set in [Appendix](#)). We used fraction pairs with a numerical distance of one between the non-constant components (e.g., 2/5 vs. 2/6 and 5/7 vs. 6/8), because it has been established that the closer the numerical values are, the more difficult the judgment ([Dehaene, 1992](#); [Moyer & Landauer, 1967](#)).

In the stimulus pairs we selected, the fraction with the larger numerator on IC trials was always the correct response; therefore, if participants made a decision based solely on the numerator, their responses would always be correct. However, there was no evidence in either the prior study (Ischebeck et al., 2009) or ours that this was an actual confound, as the behavioral results suggest, and eye gaze data confirm, that participants considered both numerators and denominators on IC trials.

As mentioned previously, it has been established that the closer together two magnitudes are, the more difficult it is to select which is greater (Moyer & Landauer, 1967). This effect applies to holistic magnitudes of fractions as well as to their components, particularly when the task promotes a holistic mental representation (e.g., Faulkenberry et al., 2015; Meert et al., 2010b). Due to the selection criteria for these stimulus pairs, the average difference in magnitudes between fractions varies with task condition, making condition and magnitude difference collinear in all regression models. In particular, IC was the most difficult condition due to the structure of the numerical relationships, but could also have been difficult because it had smaller magnitude differences between the pairs than did the other conditions. Although magnitude difference is statistically inseparable from effects of condition within this stimulus set, the performance and gaze behavior exhibited by participants is better explained by condition differences. Thus, we conducted our investigation with a focus on condition instead of magnitude difference, and make suggestions in the Discussion regarding paradigm revisions for future research.

## Metrics

From the Tobii output file we calculated trial accuracy and response times (RTs), as well as the number of saccades between digits per trial (saccades/trial). We defined an area of interest (AOI) for each digit on the screen, and measured saccades through the four AOIs. Five types of saccades were possible between each of the AOIs: numerator to numerator (NN), denominator to denominator (DD), numerator to denominator (or vice versa) on the left side (NDL), numerator to denominator (or vice versa) on the right side (NDR), and saccades between one numerator and the opposite denominator (NDX; Figure 2). Saccades that originated or terminated outside of one of these AOIs were not counted.

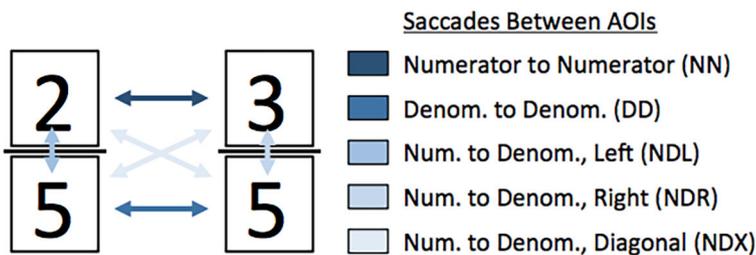


Figure 2. Examples of AOI saccades between the four AOIs, which are shown here as rectangles around the numbers.

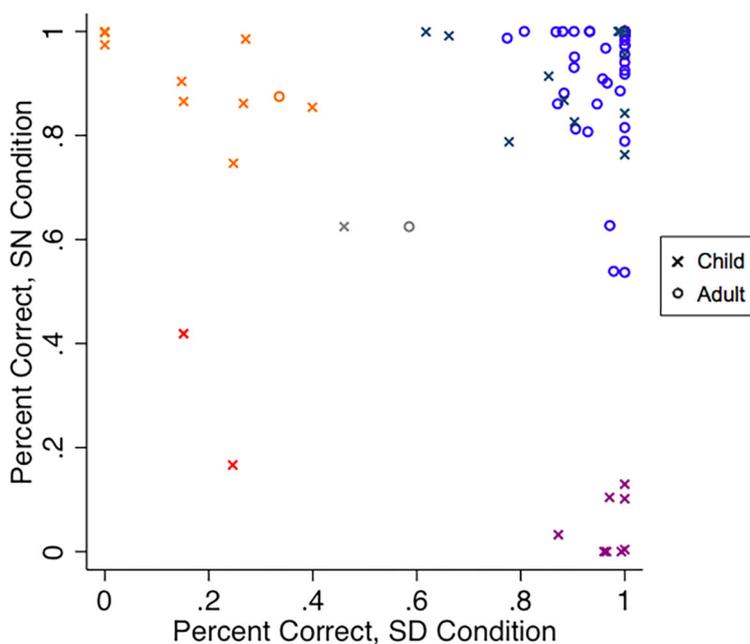
## Data Selection

Saccades between AOIs were defined by the consecutive changes in fixation recorded by the eye tracker between our four AOIs. Typical eye fixations last from 100–500 milliseconds (Holmqvist et al., 2011). For the majority of samples, data from both eyes were available and were averaged to determine gaze location; however, a valid recording from one eye is sufficient for the Tobii software to determine which AOI the participant was fixating. Any set of samples within a single AOI that lasted less than 40 milliseconds we interpreted to be a

transit between AOIs instead of a true fixation and thus were dropped. Any contiguous samples in the same AOI that were separated by fewer than 300 milliseconds of missing samples were concatenated, under the assumption that the disruption was caused by a blink.

The adults had an overall average accuracy of 91%, varying across conditions as follows: SD 95% ( $SD = 22\%$ ), SN 91% ( $SD = 29\%$ ), CO 94% ( $SD = 23\%$ ), and IC 80% ( $SD = 40\%$ ). The children had an overall average of 66% with a similar range of accuracies across conditions: SD 67% ( $SD = 47\%$ ), SN 66% ( $SD = 47\%$ ), CO 75% ( $SD = 44\%$ ), and IC 49% ( $SD = 50\%$ ). The children's poor performance was unsurprising, given they were just learning fractions at the time of the experiment, but the very large standard deviations in the child group prompted us to look more closely at the performance distribution.

Plotting average accuracy on the SN condition against the SD condition (Figure 3) revealed distinct patterns of performance.



*Figure 3.* Individual participants' average accuracy on the SN condition plotted over their accuracy on the SD condition, colored by groups resulting from a hierarchical clustering algorithm. Children are denoted as Xs and adults as Os.

*Note.* The grouping of children and adults in the top right (blue) responded consistently correctly on both SN and SD conditions. The groups of mostly children in the top left (orange) and lower right (purple) corners responded correctly on only one of those conditions, indicating they are operating on simplistic heuristics. Participants responding at chance (gray) or low on both conditions (red) were excluded, as well as the adult who was not clustered into the two-rule group.

Nearly all adults and a large subset of children had high accuracy scores on both SD and SN, indicating that they knew and could appropriately apply both the larger-numerator and smaller-denominator rules. However, two other subsets of participants had high accuracy scores on one condition and low scores on the other, indicating that they applied only one of those rules to all trials. A participant who consistently selects the larger number will respond correctly on all SD trials, for which the larger-numerator rule applies, and will respond incorrectly on all the SN trials, for which the correct response is the fraction with the smaller denominator. By contrast, a participant who consistently selects the smaller number will respond correctly on SN trials and

incorrectly on SD trials. To illustrate this distinction, consider the sample problems in Figure 1. A participant operating on a large-number bias would correctly select  $4/7$  as larger than  $3/7$ , but would incorrectly choose  $3/5$  as larger than  $3/4$ —that is, would perform well on SD trials but poorly on SN trials. A participant operating on a small-number bias would correctly select  $3/4$  as larger than  $3/5$ , but would incorrectly select  $3/7$  as larger than  $4/7$ , thereby performing poorly on SD trials but well on SN trials. Both of these biases display an incomplete understanding of the fraction rules.

A clustering algorithm including all subjects confirmed these sub-groupings. We separated the child group into those who applied two rules and those who applied only one, regardless of which rule they applied. Three children performed at or below chance on both SD and SN conditions and were not clustered with either the one-rule or two-rule groups; therefore, they were excluded. One adult participant was clustered with a one-rule group, and another fell outside the rule clusters, so they were also excluded.

Rinne et al. (2017) previously used a latent clustering algorithm that identified a group of 4<sup>th</sup>-6<sup>th</sup> grade learners who consistently selected the fraction that contained the largest number, regardless of whether that number was in the numerator or denominator. Rinne et al. posited that the heuristic of selecting the larger number demonstrates no understanding of fractions, whereas a partial understanding of fractions was exhibited by a distinct group of learners who consistently selected the smaller number. Learners often transitioned from the large-number heuristic to the small-number heuristic, and rarely the other way, suggesting that the small-number heuristic serves as a waypoint as learners develop normative understandings. Although Rinne et al. found that the small-number heuristic seemed somewhat more sophisticated than the naïve large-number heuristic, our sample was not large enough to test those subgroups separately, and so we combined them into a group that we call one-rule children. The final groups were comprised of an adult group of 36 participants, a one-rule group of 17 children, and a two-rule group of 12 children (Table 1).

Table 1

*Mean and Standard Deviation of Accuracy Measures for Each Condition, plus Demographic Data, by Sample*

Condition	One-Rule Children ( $n = 17$ )		Two-Rule Children ( $n = 12$ )		Adults ( $n = 36$ )	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Same Denominator (SD)	0.55	0.50	0.92	0.28	0.97	0.22
Same Numerator (SN)	0.51	0.50	0.94	0.25	0.92	0.26
Congruent (CO)	0.64	0.48	0.91	0.29	0.94	0.23
Incongruent (IC)	0.57	0.50	0.37	0.49	0.80	0.39
Age	10.60	0.52	10.63	0.67	20.36	1.17
% Female	36.00		65.00		67.00	

## Analyses

To accommodate the presence of the one-rule group of children, we modified our analytic plan to test for differences in eye movement behavior on specific conditions that were accessible to all groups. First, we validated our supposition that adults would be more efficient than children by testing for differences in RTs and total number of saccades. Next, we tested for differences among all groups in percent of relevant saccades, specifically on the SD and SN conditions. Saccades between numerators (NN) are relevant for the SD condition, and saccades between denominators (DD) are relevant for the SN condition. Because we combined the

one-rule groups who were consistently correct on either SD or SN, we tested for group differences in the percent of saccades on a given trial that were relevant for the problem (i.e., NN saccades for the SD condition, and DD saccades for the SN condition). Finally, we tested for differences between the two-rule children and adults on all types of saccades in the CO and IC conditions, excluding the one-rule children for whom these conditions were too difficult. In the CO and IC conditions all types of saccades could be relevant, depending on one's comparison strategy, and so we investigated whether a particular pattern of saccades was more prevalent for one group or the other.

All analyses were executed as mixed models with a random effect of subject. In each analysis, the addition of the subject factor resulted in a highly significant likelihood-ratio test over a base model that included no predictor variables. Thus, we additionally ran mixed models controlling for subject dependency and testing for one or more effects of condition, group, accuracy, or saccade types.

## Results

### Group Differences in Task Efficiency: RTs and Total Number of Saccades

Accuracy results are reported above, as they were used to define participant groups; here, we report on RT and eye gaze data. To confirm that adults performed more efficiently than children on this task, we conducted two mixed regressions with mean RTs and total number of saccades per trial as the outcome variables. After establishing significant participant-level dependence as captured by a random effect of subject, we added the categorical variables of task condition and group to each analysis.

With respect to RTs, the adults did indeed respond more quickly than the children (1-rule:  $z = 2.22$ ,  $p = .026$ ; 2-rule:  $z = 2.56$ ,  $p = .011$ ;  $f^2_{\text{group}} = 0.01$ ), although the effect sizes for group were weak, and there was no difference between the two groups of children on RTs (Figure 4a). Thus, adults responded slightly more quickly than both the 1-rule and 2-rule children, who did not differ from each other.

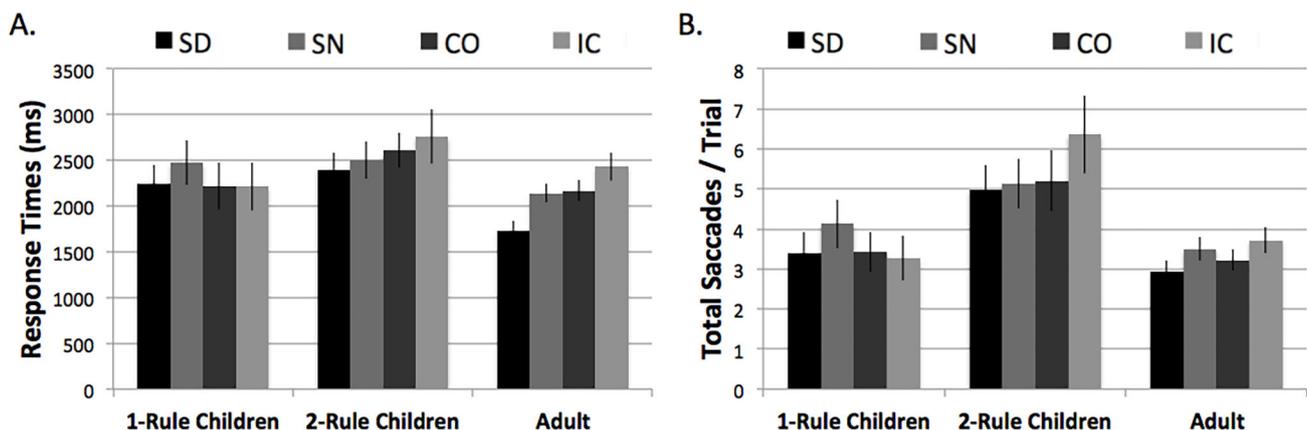


Figure 4. (a) Response times and (b) total number of saccades of interest per trial for each condition, separated into the final groups of one-rule children, two-rule children, and adults. Error bars represent standard errors.

Using SD as the reference condition, all groups responded more slowly on SN, CO, and IC than on SD (SN:  $z = 4.55$ ,  $p < .001$ ; CO:  $z = 4.78$ ,  $p < .001$ ; IC:  $z = 7.82$ ,  $p < .001$ ;  $f^2_{\text{condition}} = 0.04$ ). However, there were significant

group by condition interactions of the one-rule group with both CO and IC (one-rule by CO:  $z = -2.79$ ,  $p = .005$ ; one-rule by IC:  $z = -4.61$ ,  $p < .001$ ), showing that those children did not exhibit the same slowing down on the more difficult conditions that the two-rule children and the adults did. Note that we calculated effect sizes according to the method given in Selya, Rose, Dierker, Hedeker, and Mermelstein (2012) which does not allow for estimation of effect sizes of both main and interaction effects, and so we report  $f^2$  only for main effects and point out interactions where they added explanatory value to the regression model. In general, these results indicate that adults were indeed more efficient at making numerical judgments than children, and that both adults and two-rule children, but not one-rule children, were responsive to the increasing levels of task difficulty.

With respect to the eye-tracking data, the pattern observed for the total number of saccades of interest (i.e., those between AOIs) per trial was not redundant with that observed for RTs (Figure 4b). Instead, the two-rule children made significantly more saccades on all conditions than either the one-rule children or the adults (1-rule:  $z = -2.16$ ,  $p = .031$ ; adults:  $z = -3.16$ ,  $p = .002$ ;  $f^2_{\text{group}} = 0.01$ ), and especially on the IC condition as compared to the one-rule group (1-rule:  $z = -3.45$ ,  $p = .001$ ; adults:  $z = -1.52$ ,  $p > .05$ ), although the effect size was weak. Both the two-rule children and adults made more saccades on the most difficult condition, IC, than on the easiest, SD ( $z = 3.93$ ,  $p < .001$ ;  $f^2_{\text{condition}} = 0.025$ ), whereas the one-rule children did not ( $z = -2.68$ ,  $p = .007$ ); this result parallels the RT pattern. The adults also made more saccades on SN than SD ( $z = 2.83$ ,  $p = .005$ ), while neither the one-rule or two-rule groups did. Note that this metric includes only saccades between AOIs: it excludes all saccades originating or terminating in an area of the screen that lies outside an AOI (see Figure 2). This saccades metric indicates that adults were more sensitive to the varying difficulty levels of conditions than the children, and that the two-rule children made more eye movements in all conditions than either the one-rule children or the adults.

In summary, the adults differed from children in their overall faster RTs, and in their saccade sensitivity between SD and SN conditions. The two-rule children differed from their one-rule peers and from adults by making more saccades on all conditions. The one-rule children were distinguished by their lack of RT sensitivity to condition difficulty.

## Group Differences in Saccades on SD and SN

Next, we tested for qualitative differences in gaze behavior that would indicate whether the problem-solving strategies of novices differed from those of experienced adults. For this analysis, we focused on the easier conditions: the SD and SN trials. Because we had created the one-rule group by combining the children who consistently selected large numbers with those that consistently selected small numbers (i.e., those who used only one rule or the other), we collapsed the SD and SN conditions and created a new metric that would apply to both conditions. For both SD and SN, only one type of saccade is relevant (NN for SD and DD for SN; Figures 1 and 2). Thus, we created a metric of the percentage of relevant to total number of saccades between AOIs per trial (Figure 5) and tested for differences between all three groups. We conducted a mixed regression on only correct trials with a random effect of subject and categorical predictor variables of group and condition. Note that the condition variable tests for differences within subjects for the two-rule and adult groups, as those participants generally answered correctly on both SD and SN. However, the condition variable tests for differences between sub-groups of the one-rule group, because some participants answered correctly on SD and others answered correctly on SN. Thus, the condition factor is difficult to interpret and was included solely as a control variable, to clarify the interpretation of any effects of group or accuracy.

NN saccades were by far the most prevalent type of saccade for both SN and SD correct trials, for all three groups; on SD trials the NN saccades comprised the “relevant” metric, while looking between numerators on SN trials provided only redundant information. On SD trials, 48% of adults’ saccades were between the two relevant numbers (Figure 5); similarly, 55% of two-rule children’s saccades and 56% of one-rule children’s saccades were between the relevant numbers. On SN trials, 18% of adults’ saccades, 22% of two-rule children’s saccades, and 29% of one-rule children’s saccades were between the relevant numbers (i.e., DD saccades). Adults exhibited a numerically smaller percentage of relevant saccades than both groups of children on both conditions, but only the difference between the adults and the one-rule children reached the statistical threshold ( $z = -2.42, p = .016; f^2_{\text{group}} = 0.003$ ), with a weak effect. The difference between the two-rule children and the other groups did not reach statistical threshold ( $z_{\text{one-rule}} = 0.91, p = .36; z_{\text{adults}} = -1.32, p = .19$ ); this group fell between the one-rule children and the adults. The effect size of condition was much larger than that of group because all groups made a higher percentage of relevant saccades on correct SD trials than on correct SN trials ( $z = -13.15, p < .001; f^2_{\text{condition}} = 0.23$ ). As noted above, however, condition and sub-group were confounded within the group of one-rule children, because some children were correct on SD and others correct on SN, so it is difficult to make a general interpretation for that group. Overall, the groups exhibited a similar pattern of making a large percentage of relevant saccades on the SD condition and fewer relevant saccades on the SN condition, with the one-rule children making the highest percentage of relevant saccades and the adults making the lowest.

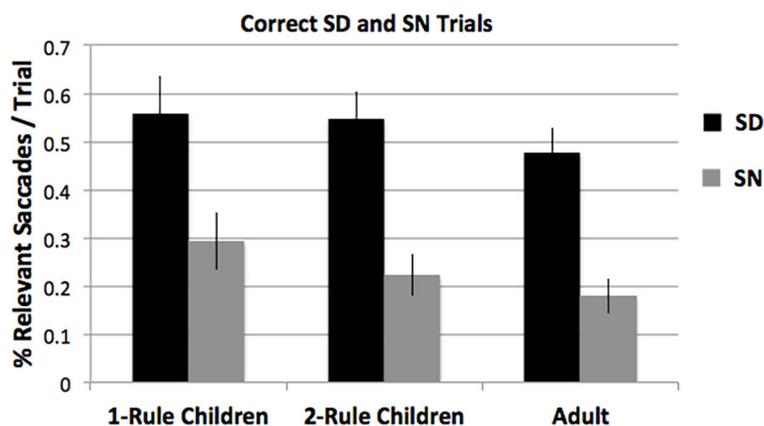


Figure 5. Mean percentage of relevant saccades per correct trial for the SD and SN conditions, separated by group.

As mentioned above, our planned analyses did not account for the unexpected difference in children’s behavior, as revealed by the accuracy profiles that showed a substantial number of children operated with either a large-number or small-number bias. The large-number bias children responded correctly to the SD trials (e.g., indicating that 4/7 is greater than 3/7) and incorrectly to the SN trials (e.g., indicating that 3/5 is greater than 3/4), and the small-number bias children responded correctly on SN trials (e.g., 3/4 is greater than 3/5) and incorrectly on SD trials (e.g., 3/7 is greater than 4/7). To explore the gaze behavior of these subgroups, we created a metric of percentage of redundant saccades per trial, comprised of saccades between identical numbers as a percentage of total saccades per trial (i.e., the percent of saccades between numerators in the SN condition and between denominators in the SD condition). Because some saccades in a trial were vertical or diagonal, the percentages of relevant and redundant saccades were not complementary. For this exploration we chose to include both correct and incorrect trials because all participants, even those in the one-rule group,

behaved generally consistently within conditions; therefore, their incorrect responses might show what gaze behavior predicated their mistaken reasoning.

This exploratory analysis tested for differences between relevant and redundant saccades across and within three groups: two-rule children who appropriately applied both large-number and small-number rules, one-rule children who exhibited a small-number bias, and one-rule children who exhibited a large-number bias. In the SD condition, all groups made more relevant than redundant saccades ( $z_{\text{redundant}} = -14.33$ ,  $p < .001$ ; all  $p_{\text{group}} > .3$ ), mirroring the main analysis described above. In the SN condition, however, the groups exhibited distinct gaze behavior, indicated by significant group by saccade-type interactions so we report here those contrasts of relevant to redundant saccades within groups during SN trials. The two-rule children made approximately equal numbers of relevant and redundant saccades during SN trials ( $z = 0.48$ ,  $p = .63$ ). The small-number subgroup made more relevant than redundant saccades on SN trials ( $z = 2.61$ ,  $p = .009$ ), examining the denominators more than the numerators, while the large-number subgroup made more redundant than relevant saccades on these trials ( $z > 4$ ,  $p < .001$ ). Although this exploratory analysis was underpowered, it suggests that there is a meaningful difference between the children who exhibit a large-number bias compared to those who exhibit a small-number bias, and warrants further investigation.

## Group Differences in Saccades on CO and IC

The CO condition could be solved by operating on either the larger-numerator rule or the smaller-denominator rule, and thus accuracy was generally very high for this condition (Table 1). The IC condition, however, set those two rules in conflict, such that participants needed a different strategy in order to select the larger fraction. Accordingly, accuracy among the children's groups was very low for IC. Given that one-rule children—by definition—had not mastered the basics of fractions, performing poorly when comparing fractions with shared components, we reasoned that their performance on trials with no shared components would be uninterpretable. However, we posited that the two-rule children, despite performing poorly on IC trials, might demonstrate gaze behavior that illuminates the challenges faced by novices when attempting to integrate multiple rules. Therefore, in the following analyses we tested only the two-rule children and the adults, and included both correct and incorrect trials because there were too few correct trials on IC to test.

In the CO and IC conditions, all saccades between numbers are relevant, depending on the selected strategy, and many strategies are appropriate. Therefore, we tested the percentage of each type of saccade separately (i.e., NN, DD, NDL, NDR, NDX). Because many trials contained none of the target saccades, and those zero values were included in the calculations and in Figure 6, the overall averages are quite low (see Discussion for our interpretation). As previously, we conducted mixed regression analyses with a random effect of subject, and set the percentage of each type of saccade as a separate outcome measure (Table 2). The only metric that displayed a difference between two-rule children and adults was the per-trial percentage of NDR saccades, showing that children made more eye movements between numerators and denominators on the right side of the screen than did adults. This difference surpassed the Bonferroni adjusted alpha level of .01 for IC ( $z = -2.83$ ,  $p = .005$ ;  $f^2_{\text{group}} = 0.001$ ) but not CO ( $z = -1.78$ ,  $p = .08$ ;  $f^2_{\text{group}} < 0.001$ ), although Figure 6 shows that this distinction is only a matter of degree, and both effects are very weak. For all other metrics, percentages of NN, DD, NDL and NDX saccades per trial, the two groups were not appreciably different. Overall, although all numerical relationships are relevant for CO and IC trials, the children focused more on the relationships between numerator and denominator than did the adults.

Table 2

Regression of Gaze Metrics on Group, Comparing Adults to Two-Rule Children

Percentage of Saccades per Trial	Congruent (CO)			Incongruent (IC)		
	$\beta$	SE	z	$\beta$	SE	z
Numerator-Numerator (NN)	.063	0.098	0.64	.038	0.095	0.40
Denominator-Denominator (DD)	.004	0.062	0.06	.018	0.066	0.27
Numerator-Denominator Left (NDL)	-.017	0.034	-0.51	-.010	0.036	-0.29
Numerator-Denominator Right (NDR)	-.065	0.037	-1.78	-.089**	0.032	-2.83
Numerator-Denominator Cross (NDX)	.020	0.037	0.55	.046	0.039	1.20

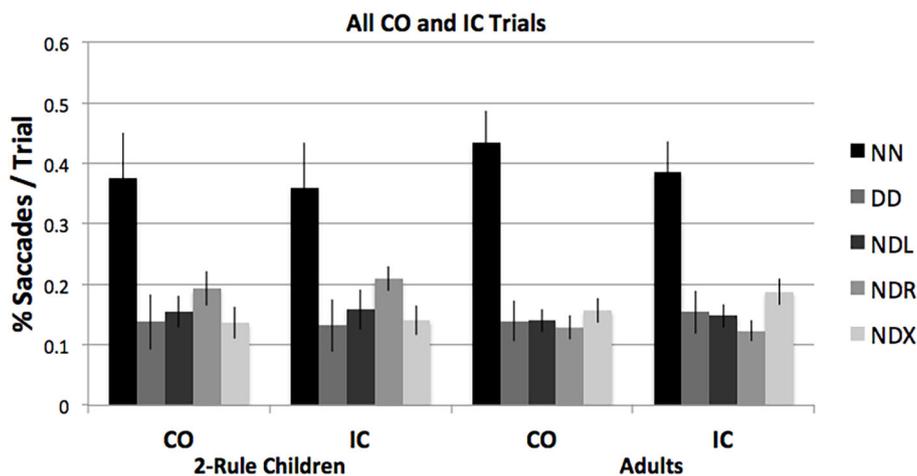
\*\* $p = .005$ .

Figure 6. Average percentages of each type of saccade per trial, for two-rule children and adults on all correct and incorrect Congruent (CO) and Incongruent (IC) trials.

## Discussion

In this study we sought to identify the strategies that support mathematical reasoning, and thereby point to potential instructional tools for new learners. To this end, we investigated how children who are beginning to learn fractions solve a fraction task, as compared with adults. We used the fraction comparison task as the setting for inquiry, because successful behavior on this task has been established in adults but not yet characterized in children, and because the task is displayed in such a way that eye-tracking methodology can provide insight into the form of relational reasoning that participants engage in during the task. In addition to having greater familiarity with the mathematical rules that govern the task, adults have higher levels of supporting cognitive skills that are likely to increase their task efficiency. To identify the strategies that are associated with successful mathematical reasoning, we measured the raw numbers and percentages of different types of eye movements made by children and adults as they made mathematical comparisons.

Considering the task as a whole, adults demonstrated greater efficiency than children, both responding more quickly and making fewer eye movements around the screen. This result is not surprising, as adults have quicker cognitive processing speed than children (Kail, Lervåg, & Hulme, 2016; Kail & Salthouse, 1994) and are more experienced with the type of mathematical reasoning elicited by this task. Furthermore, adults have

higher levels of working memory than do children (Gathercole, 1999) which may have allowed them to encode the numbers with fewer eye movements than the children needed.

Of the four conditions in the task, two required only a single comparison between either numerators or denominators. We took high accuracy on both of these conditions as an indicator that participants were familiar with both of the following rules: 1) given equal denominators, the larger fraction is the one with a larger numerator, and 2) given equal numerators, the larger fraction is the one with the smaller denominator. Almost all of the adults and 12 of the 29 children performed with high accuracy on both of the same-component conditions. The remaining children consistently answered in accordance with only one of the two rules, thereby performing well on one of the associated task conditions and poorly on the condition associated with the other, unknown or neglected, rule. Therefore, we split the group of children into those who responded in accordance with two rules and those who responded in accordance with one rule, and tested for qualitative and quantitative differences between these one-rule and two-rule children.

We found that the two groups of children exhibited quantitative differences on both RT and total number of saccades of interest made per trial: the two-rule children, who performed more accurately overall, did so by taking more time to respond and making more saccades between numbers. Interestingly, the difference between one-rule and two-rule groups was more exaggerated in the total saccades metric than in RTs, indicating a difference in gaze behavior that was not detected in terms of overall RTs. Specifically, the two-rule group made far more saccades between numbers than either the one-rule group or the adults, disproportionate to the difference in RTs. This pattern may indicate that two-rule participants focused more on the numerical relationships and therefore made disproportionately more eye movements between numbers than their RTs would predict. This would be interesting to investigate further with additional participants and additional metrics.

Another difference between the groups is that the children who responded in accordance with both rules exhibited slower RTs and a greater number of saccades per trial for the most difficult condition, as did the adults, whereas the children who operated on only one rule did not seem to be affected by the increased task difficulty. We interpret the faster RTs of the less knowledgeable group as a lack of persistence when faced with a challenge beyond their knowledge. The two-rule group also exhibited very low accuracy on this most difficult condition, suggesting it was beyond their knowledge also, yet their slow RTs and high number of saccades indicate they persisted in their attempts. Educators currently identify persistence or lack thereof in general classroom behavior; as computerized assessments are becoming more widely used by teachers, RT data would allow them to identify persistence on a trial level and therefore better discern which types of challenges promote productive struggle, versus those that are beyond reach, for individual learners.

Turning to our primary question of interest, we tested for differences in gaze patterns, that is, the relative prevalence of different types of saccades that would indicate different problem-solving strategies. We had expected that adults' expertise would lead to distinct strategies—both from the children and between conditions—which could be informative for instructors. Instead, we found that when participants responded correctly, their gaze patterns looked very similar to each other, regardless of age or proficiency. Specifically, despite the large quantitative differences between the one-rule and two-rule children, their percentages of different types of saccades were the same on correct SD and SN trials. Thus, when they knew and applied the correct rule, their eye movements aligned with the normative strategy of comparing the relevant numbers and looking relatively less at the redundant numbers. Adults exhibited this pattern as well, although to a lesser degree, likely because

they made far fewer saccades overall. Therefore, once a rule was learned, novices and adults applied it in the same way.

However, when participants responded incorrectly, or when the task demands exceeded their knowledge base, their confusion was marked by relatively more saccades toward redundant or unnecessary information. All participants made more irrelevant saccades during SN trials than they did during SD trials—and for some participants, redundant saccades surpassed relevant saccades during the SN trials. Our exploratory analysis showed that the large-number bias subgroup of children made more redundant than relevant saccades during the SN trials, and the two-rule children made approximately equal percentages of relevant and redundant saccades on these trials. Protracted focus on the equal numerators suggests confusion on how to evaluate them, and is not helpful as there is no information to be gleaned once the equality is encoded.

An interesting exception from the SD and SN exploratory analysis is the small-number bias children, who consistently selected the fraction with the smaller number regardless of whether that number was in the numerator or denominator position. Like the other groups, they exhibited more relevant than redundant saccades on the SD condition, but despite their normative gaze behavior, they largely selected the *incorrect* response. Unlike the other groups, however, they made more relevant than redundant saccades on the SN condition, in which they performed very well. The fact that they are consistently looking at the most helpful information, and yet sometimes reasoning incorrectly about it, supports the well-established idea that reconciling different rules about fractions is conceptually challenging, yet also provides insights as to how new learners approach that conceptual challenge.

Rinne et al. (2017) identified children with a small-number bias as having a more sophisticated understanding of fractions than those operating on a large-number bias. Our findings extend their conclusion by revealing the distinct problem-solving approaches of these groups. While the less-sophisticated large-number bias group attended to the relevant information only in the cases that were accessible to them (that is, on SD but not SN trials), the small-number bias group attended to the relevant information in both conditions, even when they ultimately made the incorrect selection. Thus, honing one's attention may be the precursor to building reasoning skills that undergird conceptual growth.

The gaze patterns of the two-rule children provided a similar indicator of misdirected attention on the more difficult conditions. Although the two-rule children knew and could apply both the larger-numerator and smaller-denominator rules in the easier conditions, the mixed pair conditions presented an additional challenge. For the CO pairs, they could follow either rule and arrive at the correct decision, but the IC pairs required integration of the rules or application of a specific strategy. Integrating multiple numerical sets is both mathematically and relationally difficult; accordingly, both adults and two-rule children performed well on the CO condition and poorly on the IC condition.

On the IC condition, where they performed most poorly, the two-rule children made more saccades between the numerator and denominator in the right fraction than did adults. They exhibited similar behavior on the CO condition, but the group difference only reached statistical significance on the IC test. Because a number of strategies would be successful in the mixed pair case, saccades between numerators and denominators are indeed relevant, and corroborate prior studies that show people make a greater number of vertical saccades during mixed pair trials (Obersteiner & Tumpek, 2016). It is thought that vertical saccades indicate an attempt to integrate the two values into an estimated (or calculated) magnitude for the fraction. Thus, these data could

be interpreted as evidence that the two-rule children attend to the information that may help them make the conceptual leap to fractions as integrated magnitudes.

However, to accurately make a comparison it is necessary to assess the integrated magnitude of both left and right fractions; yet, the two-rule children looked preferentially to the fraction on the right during the IC condition. Failure to attend to relevant information may indicate that these trials were beyond their reach. An alternative explanation is that if participants look first to the left side of the screen, the left fraction would exhibit a primacy effect. Then, the working memory constraints of children would lead them to look more frequently at the right fraction to help them encode it after their working memory has reached capacity. Adults' working memory is likely sufficient to encode all numbers on one scan and they do not need to make repeated saccades to either fraction for the purpose of encoding. This supposition could be evaluated with a scan path analysis, which we did not have the power to undertake here.

Alternatively, these two findings taken together—that participants looked more frequently at less-informative areas of the screen when they were unsure of the appropriate problem-solving strategy—may reflect the difficulty associated with integrating numerical relationships. In this study, the fraction comparison task was novel for the children, and their eye movements made apparent the relationships that were challenging for them: equal numerators and the numerator-denominator relationship in the case of mixed pairs. For the shared-component trials, the larger-smaller relationship is apparent, but integrating that with an equal relationship, particularly in the case of equal numerators, is conceptually challenging. For the mixed pair trials, participants made more vertical saccades, perhaps attempting to integrate the numerator and denominator into a magnitude, which is conceptually even more difficult.

Importantly, these findings are richer for the use of eye-tracking methodology, which provided insights beyond the traditional behavioral metrics of RT and accuracy. In particular, participants tended to pay more attention to redundant information on trials that were well beyond their conceptual reach. However, attention to relevant information may indicate that participants were ready to approach the next conceptual challenge, even if they responded incorrectly on those trials, as in the cases of two-rule children on the IC condition and the small-number bias subgroup on SD and SN trials. Additionally, the children who were able to switch between fraction rules (i.e., they selected the fraction with the larger numerator or the smaller denominator) made a greater overall number of eye movements than did the one-rule children or the adults, out of proportion to the additional time they spent on the problems. These findings are novel in the literature.

One important caveat is that we found a lower number of saccades per trial than did other researchers: our participants averaged three to five saccades of interest per trial, while the participants in Ischebeck, Weilharter, and Körner's study averaged 6–9 saccades per trial, and those in Obersteiner and Tumpek's (2016) study averaged 7–12 saccades per trial depending on the type of fraction pair. There are three plausible, non-mutually exclusive, explanations for this discrepancy. First, participants in our study responded much more quickly than participants in other studies. This is likely because we opted to keep the numbers small and the trials accessible to young children, and thus the problems may have been too easy for adults. Obersteiner and Tumpek used only two-digit numbers, which made the problems more difficult for adults, and thus they spent more time and made more saccades per trial. However, even our child participants responded more quickly than the adults in other studies; it may also be the case that our verbal instructions to answer quickly created an experimental environment that differed from the other studies. A second plausible reason for the lower

number of saccades is that we counted only those that originated or terminated within a defined space around the numbers, whereas other researchers made less conservative analytical choices.

Finally, a third possible reason for the lower number of saccades in our study than in other studies is that we selected a screen layout that maintained the visual familiarity of fractions, for the sake of the new learners. [Ischebeck et al. \(2016\)](#), by contrast, promoted higher numbers of saccades by adding visual noise around the numbers so as to prevent participants from encoding the numerals without fixating directly on the numbers. Our decision to maintain the familiar fractions format may have made it possible to use peripheral vision.

If people used peripheral vision, it may explain another disparity with previously-published findings. [Huber et al. \(2014\)](#) found that adults spent more time on denominators than numerators, whereas our adults did not show that preference. Instead, in our data, participants looked preferentially between numerators on all conditions. We interpreted the focus on numerators as a carryover practice of reading top to bottom. Indeed, Obersteiner and Tumpek also found a greater number of fixations on numerators than denominators, except when the fractions shared identical numerators. However, Ischebeck, Weilharter, and Körner conducted a scan path analysis which indicated that people first “read” the left fraction and then the right, but do not necessarily make saccades between numerators within their initial scans. Therefore, the prevalence of NN saccades in our results may be due to peripheral vision.

Future research using this paradigm should continue to address the problem of peripheral vision. We chose to design the screen to put the numbers in proximity of the vinculum so that they were easily recognizable as fractions, but doing so may have weakened our analyses. Other researchers have used visual noise or greater distance between numbers to encourage eye movements, study design choices that work well for adult participants, but may have challenged children’s interpretation of the numbers as fractions.

Additionally, future research using this paradigm should adjust the stimulus set such that each condition contains the same range of magnitude differences between fraction pairs. In this set, the most difficult condition also had the smallest magnitude differences, and thus condition and magnitude difference were confounded. Because our children were struggling to understand the concept of fractions as an integrated magnitude, we considered it unlikely that their behavior was impacted by the overall magnitude difference between fractions, and thus we interpreted our data in the context of conditions. Additional studies could clarify the findings by adjusting the stimuli.

One important question regarding this task to be addressed in future research is how to best support children who are struggling with acquiring the basic rules. In this study we grouped them as one-rule children because of our limited sample size, but [Rinne et al. \(2017\)](#) found that the children who exhibited a small-number bias were more advanced than the children who exhibited a large-number bias. The wider variation in accuracy within our small-number bias group supported this: the children who exhibited a small-number bias nevertheless responded correctly on some of the SD trials, whereas the children who exhibited a large-number bias did so consistently—to the point of getting almost none of the SN trials correct. Our exploratory analysis comparing these subgroups also corroborated this ranking by showing that the small-number bias children looked at the relevant numerical relationships even when they responded incorrectly, whereas the large-number bias children did not. A larger sample of these one-rule children may be able to detect meaningful gaze differences between these groups and thereby provide additional insights to educators as they introduce these difficult fractions concepts.

A larger sample of children would also enable researchers to regard the two-rule children—that is, the ones who had successfully acquired at least the basic concepts of fractions—as the standard for learning. We set adults as the standard, hoping to identify gaze patterns associated with proficient problem-solving. However, either because this task was mathematically too simplistic for adults, or because their working memory is better, they made far fewer saccades than children. Thus, it was difficult to characterize their problem-solving strategies. Instead of comparing novices to experienced adults, future research may glean more useful insights by making additional comparisons between successful and struggling students.

Nevertheless, our findings are relevant for educators in that they point to the numerical relationships that are challenging for novices. Because understanding fractions requires attention to numerical relationships, the fact that novices are indeed attending to those relationships is heartening; yet, the children who struggled the most seemed drawn to redundant numerical relationships. The children who had correctly acquired the basic fraction concepts attended to the relevant information on the simpler trials and seemed poised to begin evaluating fraction magnitudes as defined by numerator-denominator relationships. Supporting their attention to relevant information and their relational reasoning will help children acquire normative fraction knowledge.

## Funding

This research was supported by a James S. McDonnell Scholar Award to S.A.B, and by Institute of Education Sciences predoctoral training grant R305B090026 to the University of California, Berkeley and A.T.M.S.

## Competing Interests

The authors have declared that no competing interests exist.

## Acknowledgments

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

## Author Contributions

Data were collected and analyzed by A.T.M.S. and J.L.C.

Manuscript was written and edited by A.T.M.S., J.L.C., and S.A.B.

## Data Availability

For this study a dataset is freely available (see the [Supplementary Materials](#) section).

## Supplementary Materials

The Supplemental Materials include the following items (for access see Index of [Supplementary Materials](#) below).

- Subject Data files: Raw files from Tobii eyetracker, filtered fixation lists
- Analysis files: Python script used to calculate saccades, aggregated data file that formed basis of statistical analyses

**Index of Supplementary Materials**

Miller Singley, A. T., Crawford, J. L., & Bunge, S. A. (2020). *Dataset for: "Eye gaze patterns reflect how young fraction learners approach numerical comparisons"*. PsychOpen. <https://doi.org/10.23668/psycharchives.2720>

Miller Singley, A. T., Crawford, J. L., & Bunge, S. A. (2020). *Code for: "Eye gaze patterns reflect how young fraction learners approach numerical comparisons"*. PsychOpen. <https://doi.org/10.23668/psycharchives.2721>

**References**

- Bazargani, N., Hillebrandt, H., Christoff, K., & Dumontheil, I. (2014). Developmental changes in effective connectivity associated with relational reasoning. *Human Brain Mapping, 35*(7), 3262-3276. <https://doi.org/10.1002/hbm.22400>
- Bonato, M., Fabbri, S., Umiltà, 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. <https://doi.org/10.1037/0096-1523.33.6.1410>
- Booth, J. L., & Newton, K. J. (2012). Fractions: Could they really be the gatekeeper's doorman? *Contemporary Educational Psychology, 37*(4), 247-253. <https://doi.org/10.1016/j.cedpsych.2012.07.001>
- Bucci, M. P., & Seassau, M. (2012). Saccadic eye movements in children: A developmental study. *Experimental Brain Research, 222*(1-2), 21-30. <https://doi.org/10.1007/s00221-012-3192-7>
- 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. <https://doi.org/10.1007/s10649-009-9198-9>
- Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*(11), 2037-2078. <https://doi.org/10.1016/j.neuropsychologia.2006.02.006>
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44*(1-2), 1-42. [https://doi.org/10.1016/0010-0277\(92\)90049-N](https://doi.org/10.1016/0010-0277(92)90049-N)
- Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research, 36*(12), 1827-1837. [https://doi.org/10.1016/0042-6989\(95\)00294-4](https://doi.org/10.1016/0042-6989(95)00294-4)
- Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. Stuss & R. Knight (Eds.), *Principles of frontal lobe function* (pp. 466-503). New York, NY, USA: Oxford University Press.
- Dumas, D., Alexander, P. A., & Grossnickle, E. M. (2013). Relational reasoning and its manifestations in the educational context: A systematic review of the literature. *Educational Psychology Review, 25*, 391-427. <https://doi.org/10.1007/s10648-013-9224-4>
- 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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/10.1037/xlm0000153>
- Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, *3*(11), 410-419. [https://doi.org/10.1016/S1364-6613\(99\)01388-1](https://doi.org/10.1016/S1364-6613(99)01388-1)
- Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, *23*(4), 523-552. <https://doi.org/10.1007/s10648-011-9174-7>
- Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, *24*(1), 79-132. [https://doi.org/10.1016/S0160-2896\(97\)90014-3](https://doi.org/10.1016/S0160-2896(97)90014-3)
- Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science*, *14*(5), 462-466. <https://doi.org/10.1111/1467-9280.02454>
- 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. <https://doi.org/10.1016/j.jecp.2016.12.005>
- 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. <https://doi.org/10.1017/S0140525X98001769>
- Hoffer, T. B., Venkataraman, L., Hedberg, E. C., & Shagle, S. (2007). *Final report on the National Survey of Algebra Teachers for the National Math Panel*. Washington, DC, USA: US Department of Education [conducted by the National Opinion Research Center at the University of Chicago].
- 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, United Kingdom: Oxford University Press.
- Huber, S., Moeller, K., & Nuerk, H. C. (2014). Adaptive processing of fractions—Evidence from eye-tracking. *Acta Psychologica*, *148*, 37-48. <https://doi.org/10.1016/j.actpsy.2013.12.010>
- Inagaki, K., & Hatano, G. (1987). Young children's spontaneous personification as analogy. *Child Development*, *58*(4), 1013-1020. <https://doi.org/10.2307/1130542>
- Ischebeck, A., Schocke, M., & Delazer, M. (2009). The processing and representation of fractions within the brain: An fMRI investigation. *NeuroImage*, *47*, 403-413. <https://doi.org/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. <https://doi.org/10.1080/17470218.2015.1046464>
- Johnson, E. L., Miller Singley, A. T., Peckham, A. D., Johnson, S. L., & Bunge, S. A. (2014). Task-evoked pupillometry provides a window into the development of short-term memory capacity. *Frontiers in Psychology*, *5*, Article 218. <https://doi.org/10.3389/fpsyg.2014.00218>

- Kail, R. V., Lervåg, A., & Hulme, C. (2016). Longitudinal evidence linking processing speed to the development of reasoning. *Developmental Science*, *19*(6), 1067-1074. <https://doi.org/10.1111/desc.12352>
- Kail, R. V., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica*, *86*(2-3), 199-225. [https://doi.org/10.1016/0001-6918\(94\)90003-5](https://doi.org/10.1016/0001-6918(94)90003-5)
- Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child Development*, *75*(5), 1357-1372. <https://doi.org/10.1111/j.1467-8624.2004.00745.x>
- 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. <https://doi.org/10.1080/17470210802511162>
- Meert, G., Grégoire, J., & Noël, M. P. (2010a). Comparing 5/7 and 2/9: Adults can do it by accessing the magnitude of the whole fractions. *Acta Psychologica*, *135*(3), 284-292. <https://doi.org/10.1016/j.actpsy.2010.07.014>
- Meert, G., Grégoire, J., & Noël, M. P. (2010b). Comparing the magnitude of two fractions with common components: Which representations are used by 10- and 12-year-olds? *Journal of Experimental Child Psychology*, *107*(3), 244-259. <https://doi.org/10.1016/j.jecp.2010.04.008>
- Miller Singley, A. T., & Bunge, S. A. (2014). Neurodevelopment of relational reasoning: Implications for mathematical pedagogy. *Trends in Neuroscience and Education*, *3*(2), 33-37. <https://doi.org/10.1016/j.tine.2014.03.001>
- Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*(5109), 1519-1520. <https://doi.org/10.1038/2151519a0>
- National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the national mathematics advisory panel*. Washington, DC, USA: U.S. Department of Education.
- Obersteiner, A., Moll, G., Beitch, J. T., Cui, C., Schmidt, M., Khmelivska, T., & Reiss, K. (2014). Expert mathematicians' strategies for comparing the numerical values of fractions: Evidence from eye movements. In P. Liljedahl, S. Oesterle, C. Nicol, & D. Allan (Eds.), *Proceedings of the joint meeting of PME 38 and PME-NA 36* (Vol. 4, pp. 338-345). Vancouver, Canada: PME.
- Obersteiner, A., & Tumpek, C. (2016). Measuring fraction comparison strategies with eye-tracking. *ZDM Mathematics Education*, *48*, 255-266. <https://doi.org/10.1007/s11858-015-0742-z>
- Perone, S., Simmering, V. R., & Spencer, J. P. (2011). Stronger neural dynamics capture changes in infants' visual working memory capacity over development. *Developmental Science*, *14*(6), 1379-1392. <https://doi.org/10.1111/j.1467-7687.2011.01083.x>
- 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. <https://doi.org/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. <https://doi.org/10.1177/2372732216629795>

- Richland, L. E., Holyoak, K. J., & Stigler, J. W. (2004). Analogy use in eighth-grade mathematics classrooms. *Cognition and Instruction*, 22(1), 37-60. <https://doi.org/10.1207/s1532690Xci2201>
- Richland, L. E., & McDonough, I. M. (2010). Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology*, 35(1), 28-43. <https://doi.org/10.1016/j.cedpsych.2009.09.001>
- Rinne, L. F., Ye, A., & Jordan, N. C. (2017). Development of fraction comparison strategies: A latent transition analysis. *Developmental Psychology*, 53(4), 713-730. <https://doi.org/10.1037/dev0000275>
- Schneider, M., Heine, A., Thaler, V., Torbeyns, J., De Smedt, B., Verschaffel, L., . . . Stern, E. (2008). A validation of eye movements as a measure of elementary school children's developing number sense. *Cognitive Development*, 23(3), 424-437. <https://doi.org/10.1016/j.cogdev.2008.07.002>
- Selya, A. S., Rose, J. S., Dierker, L. C., Hedeker, D., & Mermelstein, R. J. (2012). A practical guide to calculating Cohen's  $f^2$ , a measure of local effect size, from PROC MIXED. *Frontiers in Psychology*, 3, Article 111. <https://doi.org/10.3389/fpsyg.2012.00111>
- 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. <https://doi.org/10.1080/14640748608401609>
- Smith, J. P., III (1995). Competent reasoning with rational numbers. *Cognition and Instruction*, 13(1), 3-50. [https://doi.org/10.1207/s1532690xci1301\\_1](https://doi.org/10.1207/s1532690xci1301_1)
- Sprute, L., & Temple, E. (2011). Representations of fractions: Evidence for accessing the whole magnitude in adults. *Mind, Brain, and Education*, 5(1), 42-47. <https://doi.org/10.1111/j.1751-228X.2011.01109.x>
- Sullivan, J. L., Juhasz, B. J., Slattery, T. J., & Barth, H. C. (2011). Adults' number-line estimation strategies: Evidence from eye movements. *Psychonomic Bulletin & Review*, 18(3), 557-563. <https://doi.org/10.3758/s13423-011-0081-1>
- van't Noordende, J. E., van Hoogmoed, A. H., Schot, W. D., & Kroesbergen, E. H. (2016). Number line estimation strategies in children with mathematical learning difficulties measured by eye tracking. *Psychological Research*, 80(3), 368-378. <https://doi.org/10.1007/s00426-015-0736-z>
- Wendelken, C., Ferrer, E., Ghetti, S., Bailey, S. K., Cutting, L., & Bunge, S. A. (2018). Frontoparietal structural connectivity in childhood predicts development of functional connectivity and reasoning ability: A large-scale longitudinal investigation. *Journal of Neuroscience*, 37(35), 8549-8558. <https://doi.org/10.3389/fpsyg.2018.00265>
- White, C. S., Alexander, P. A., & Daugherty, M. (1998). The relationship between young children's analogical reasoning and mathematical learning. *Mathematical Cognition*, 4, 103-123. <https://doi.org/10.1080/135467998387352>

## Appendix: Stimulus Set

Table A1

List of Stimuli Used in Fraction Comparison Task With Fifth Graders and Adults

SD	$\frac{6}{8}$	$\frac{5}{8}$	$\frac{5}{8}$	$\frac{6}{8}$	$\frac{3}{8}$	$\frac{4}{8}$	$\frac{4}{8}$	$\frac{3}{8}$	$\frac{3}{7}$	$\frac{4}{7}$	$\frac{4}{7}$	$\frac{3}{7}$	$\frac{5}{7}$	$\frac{6}{7}$	$\frac{6}{7}$	$\frac{5}{7}$
SN	$\frac{3}{5}$	$\frac{3}{4}$	$\frac{3}{4}$	$\frac{3}{5}$	$\frac{3}{7}$	$\frac{3}{6}$	$\frac{3}{6}$	$\frac{3}{7}$	$\frac{2}{4}$	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{2}{4}$	$\frac{2}{7}$	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{2}{7}$
CO	$\frac{2}{7}$	$\frac{3}{6}$	$\frac{3}{6}$	$\frac{2}{7}$	$\frac{6}{7}$	$\frac{5}{8}$	$\frac{5}{8}$	$\frac{6}{7}$	$\frac{4}{7}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{4}{7}$	$\frac{2}{5}$	$\frac{3}{4}$	$\frac{3}{4}$	$\frac{2}{5}$
IC	$\frac{6}{8}$	$\frac{5}{7}$	$\frac{5}{7}$	$\frac{6}{8}$	$\frac{2}{4}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{2}{4}$	$\frac{4}{8}$	$\frac{3}{7}$	$\frac{4}{7}$	$\frac{3}{7}$	$\frac{2}{6}$	$\frac{3}{4}$	$\frac{3}{4}$	$\frac{2}{6}$

Note. SD = same denominator; SN = same numerator; CO = congruent condition; IC = incongruent condition.