Bridging Cognitive and Education Research to Gain Insights on Fractions Learning

By

Alison Taylor Miller Singley

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Psychology

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge:

Professor Silvia A. Bunge, Chair

Professor Tania Lombrozo

Professor Geoffrey B. Saxe

Fall 2016

Abstract

Bridging Cognitive and Education Research to Gain Insights on Fractions Learning

by

Alison Taylor Miller Singley

Doctor of Philosophy in Psychology

University of California, Berkeley

Professor Silvia A. Bunge, Chair

Symbolic fractions are notoriously difficult to learn, and this difficulty has been characterized in terms of cognitive, mathematical, and educational challenges. In this dissertation I present evidence from several eye-tracking studies that illuminate participants' cognitive approaches to proportional reasoning, and analyze them from the perspectives of psychological research on relational reasoning, as well as modern education research. In my first two studies, I had participants perform fraction comparison tasks while I measured their eye movements to assess how they approached various types of problems. In the first of these studies, I compared the performance and eye gaze patterns of 5th-graders, who were just beginning to learn fractions, with those of college students. I sought to test the hypothesis, based on relational complexity theory, that children have difficulty learning fractions in part because they have difficulty integrating relationships among mental representations. This effect was present in the data, however, there were additional performance decrements unrelated to relational complexity that are better explained by the development of inhibition and cognitive flexibility. Further, I found that children who did not comprehend fraction concepts, as evidenced by their performance, still exhibited similar eye movements to those who performed well, suggesting that they encoded the relevant numerical relations even though they were not able to interpret them correctly. These findings underscore the cognitive and mathematical complexities inherent in proportional reasoning. In the second fraction comparison study, I investigated the extent to which adults applied various relational integration skills to proportional reasoning problems, and whether doing so impacted their performance. I found that they performed better on trials that could be solved more easily by componential than magnitude processing. Specifically, when there was a readily-available multiplicative factor between the two fractions, they made fewer within-fraction saccades consistent with magnitude calculation – and when they made fewer of these saccades, they performed more efficiently. This work highlights the ways in which relational thinking can support proportional reasoning. In the dissertation, I place the results of these first two studies in the context of prior research on fraction understanding, and point to

possible implications for pedagogy. In a third study, I collaborated on an investigation of children's learning trajectories during the acquisition of fraction knowledge, comparing two curricular approaches. We found that individual prior knowledge, classroom environment, and the curriculum with which the students engaged all influenced their acquisition of fractions knowledge. As is evident in this dissertation, experimental psychology and educational research provide different, and equally productive, lenses with which to explore the learning of fractions. By integrating across these theories and methods of analysis, we can more fully characterize and promote the development of proportional reasoning.

Table of Contents

Acknowledgements

I owe an immense debt of gratitude to Dr. Silvia Bunge, for giving me the opportunity to switch careers and pursue academic research, and for her guidance throughout the course of my graduate studies. I would not have had the audacity to seriously consider such an idea without the enthusiastic emotional (and then financial and tactical) support of my husband, Eric Singley, and to him I owe my happiness and utmost appreciation. Many of my family members have dedicated their lives to educating others, and so served as my inspiration for contributing to this field of research.

I would like to thank my dissertation committee members, Drs. Tania Lombrozo and Geoffrey Saxe for their insightful questions and helpful advice that shaped my research and sharpened my thinking over the past years.

I would also like to thank Dr. Allyson Mackey, who took me on as a research assistant while I was considering applying to graduate schools, and who subsequently taught me a great deal about research and academic culture. I would also like to thank Belén Guerra, who became a kindred spirit in academic interests and an inspiration. Similarly, the members of the Bunge Lab and the Fellows of the Research in Cognition and Mathematics Education have shared valuable insights, professional and personal, as I grew as a researcher, particularly Yana Fandakova, Michael Vendetti, Ariel Starr, and Zdeña Op de Macks. Jeffrey Crawford, Heather Anderson and Maia Barrow provided essential assistance with several of my research projects and to them I also owe many thanks.

Introduction

Proportional reasoning: Its vaunted and feared place in mathematics

The evolution of proportional thinking

Formal systems of mathematics were designed and developed by humans, and yet many have argued that, abstracted though they are, these systems are rooted in the biological and physical human experience. Humans certainly share a rudimentary sense of magnitude with much of the animal kingdom (Brannon, 2006; Dehaene, 1997; Feigenson, Dehaene, & Spelke, 2004). Indeed, many animals as well as human infants are capable of discriminating approximate magnitudes, and their judgments get more precise with experience and education (Starr & Brannon, 2015). These rudimentary tasks implicate the intraparietal sulcus and corresponding brain regions in humans and animal models (Nieder & Dehaene, 2009), and more advanced mathematical tasks recruit the same and additional neural substrates (Ansari, 2008; Dehaene, Piazza, Pinel, & Cohen, 2003), suggesting that basic magnitude sense forms a foundation for more advanced mathematics.

Not only are there biological foundations for our mathematical skill, but Lakoff and Núñez have additionally claimed that symbolic mathematics is based on the embodied experience of manipulating items in the real world. For example, the acts of joining and separating sets form the conceptual foundations for basic arithmetic, and Boolean logic is based in humans' early understanding of object categories (2000). Further support for the argument that mathematics is a human construct, instead of an independently existing set of natural laws, comes from the history of mathematics, which describes the evolution of formalized systems (e.g., Gallistel & Gelman, 2005). Early mathematics was comprised of symbolic transcriptions of concrete arithmetic transactions; formalized systems developed through increasingly abstracted notations and more advanced applications.

One of those abstract notations developed to represent proportional reasoning, the ability to compare and evaluate multiple sets of numerical relationships. It is still considered one of the more difficult constructs in mathematics, yet paradoxically, it has long evolutionary roots and was utilized in formal mathematics long before its notation was canonized. New evidence supports the biological basis for a ratio processing system present in human infants and animal models that – like the magnitude processing system discussed above – relies on the intraparietal sulcus (Jacob, Vallentin, & Nieder, 2012; Lewis, Matthews, & Hubbard, 2015).

Furthermore, proportional reasoning has been formally used for millennia. Eudoxus gave the first known definition of ratio around 350BC (Boyer & Merzbach, 2011), although proportion was so essential to Greek geometry that the concept may well have been used prior to this time. In approximately 250BC, early forms of fractions were being transcribed in Hindu texts, although the *a/b* notation for rational numbers would not be universally adopted until the age of the printing press (Elwes, 2013), and the general acceptance of rational numbers as an essential entailment of the formal system of mathematics occurred only within the last two centuries (Gallistel & Gelman, 2005). The now formalized notation of *a/b* indicates a proportion between a specific number of parts (*a*) and a representation of a whole (*b*), that accurately captures a relationship but obscures the associated magnitude. For example, the proportions 1/2 and 345/690 are equivalent because they represent the same relationship and describe the same relative magnitude, but the integers bear no resemblance to each other. Thus, despite its evolutionarily privileged underpinnings, and its historically vaunted place in formal mathematics, the formal notation and logical entailments of proportional reasoning make it one of the more difficult concepts in mathematics (Vamvakoussi & Vosniadou, 2004).

Pedagogical challenges of proportional reasoning

The difficulty of proportional reasoning is evidenced in countless examples of conceptual failures documented among advanced students and adults. One notorious anecdote describes a marketing focus group that thought the McDonald's 1/4-lb burger was a better deal than A&W's 1/3-lb burger at the same price, because 4 is greater than 3 (E. Green, 2014). Another cites the 2004 National Assessment of Educational Progress exam in which 50% of 8th graders failed to put 2/7, 1/12, and 5/9 in the correct order by magnitude (National Council of Teachers of Mathematics, 2007). The persistence into adulthood of conceptual failures like these (e.g., DeWolf, Grounds, Bassok, & Holyoak, 2014; Stigler, Givvin, & Thompson, 2010) suggests inadequate learning beginning with the introduction of proportions, typically in mid- to upper-elementary school, likely stemming from several specific conceptual challenges faced by students in elementary grades. First, proportions generally serve as the introduction to and definition of rational numbers, which have vastly different rules than the whole numbers that young children learn to count. Whole numbers have a defined order, such that each number has exactly one predecessor and one successor. Rational numbers have infinitely many predecessors, successors, and even interstitial numbers because they are infinitely dense (Vamvakoussi & Vosniadou, 2004). Second, as mentioned above, the magnitude of rational numbers is defined by the relationship between the numerator and denominator, independent of the magnitudes of the components themselves. It is widely thought that the rules of counting numbers exacerbate the conceptual challenges posed by rational numbers – a theory referred to as the "whole number bias" (Ni & Zhou, 2005; Siegler, 2016). Furthermore, the *a/b* notation encompasses a variety of definitions, including but not limited to *a* pieces of a *b*-sized unit, *a* items in a set of *b*, and *a* items divided into *b* parts (Behr, Lesh, Post, & Silver, 1983). All of these definitional and conceptual challenges contribute to the pedagogical challenges associated with proportional reasoning.

A Cognitive Analysis of Proportional Reasoning

Relational complexity theory

A different theoretical framework offers a supplementary explanation for the difficulty of proportional reasoning. Relational complexity theory, advanced by Halford and colleagues, analyzes the cognitive processing load imposed by different mental tasks, and concludes that effectively evaluating the relationship between two proportions requires simultaneous attention to all four component numbers (Halford, Wilson, & Phillips, 1998). Because there are four components that define a single relationship, Halford and colleagues call this a quaternary relation. Each component in a relationship imposes additional processing load, and they posit that four dimensions is the maximum mental load that individuals can represent simultaneously.

Cramer, Post, and Currier describe three primary types of mathematical problems commonly posed in educational materials, and three common solution approaches for them (1993). The problem types are 1) numerical comparison, in which learners must select which of two given ratios is greater (Fig. 1A), 2) missing value, in which learners complete a proportion given an equivalent proportion and one corresponding component (Fig. 1B), and 3) qualitative comparison, which is similar to numerical comparison problems, except that no numbers are used and learners must make a choice based on descriptions of relative quantities (Fig. 1C).

The first approach Cramer and colleagues discuss for solving these problems is a 'unit rate' strategy in which one finds the multiplicative relationship between two given measurements. That rate is then applied to the other proportion's denominator to find the numerator that makes the two proportions equivalent. In Figure 1B, the problem sets up the proportion (8 carrots)/(12 grapes) = (x carrots)/(24 grapes). The unit rate is 1.5 grapes per 1 baby carrot (or 2/3 carrots per one grape), so 2/3 of 24 is the missing number of carrots. This unit rate approach can, of course, be adapted so that the numerator is the base and the denominator the calculated quantity, depending on the ease of working with the different numbers given in the problem, or on which number is missing. Either way, the key feature of the unit rate strategy is that it is calculated within one proportion and applied to the other proportion.

The second approach is the 'factor-of-change' strategy, which essentially looks at the cross-proportion relationships. In this approach, one denominator is compared to the other proportion's denominator to determine the multiplicative factor of difference between them. Then that factor is multiplied by the base proportion's numerator to find the missing value, or to compare to the target proportion's numerator. In Figure 1B, twice as many grapes fit in the larger container as in the smaller container, and so twice as many carrots must also fit. Again, either proportion can serve as the base, and either numerator or denominator can serve as the initial calculation, based on the affordances of the problem and the learner's cognitive flexibility.

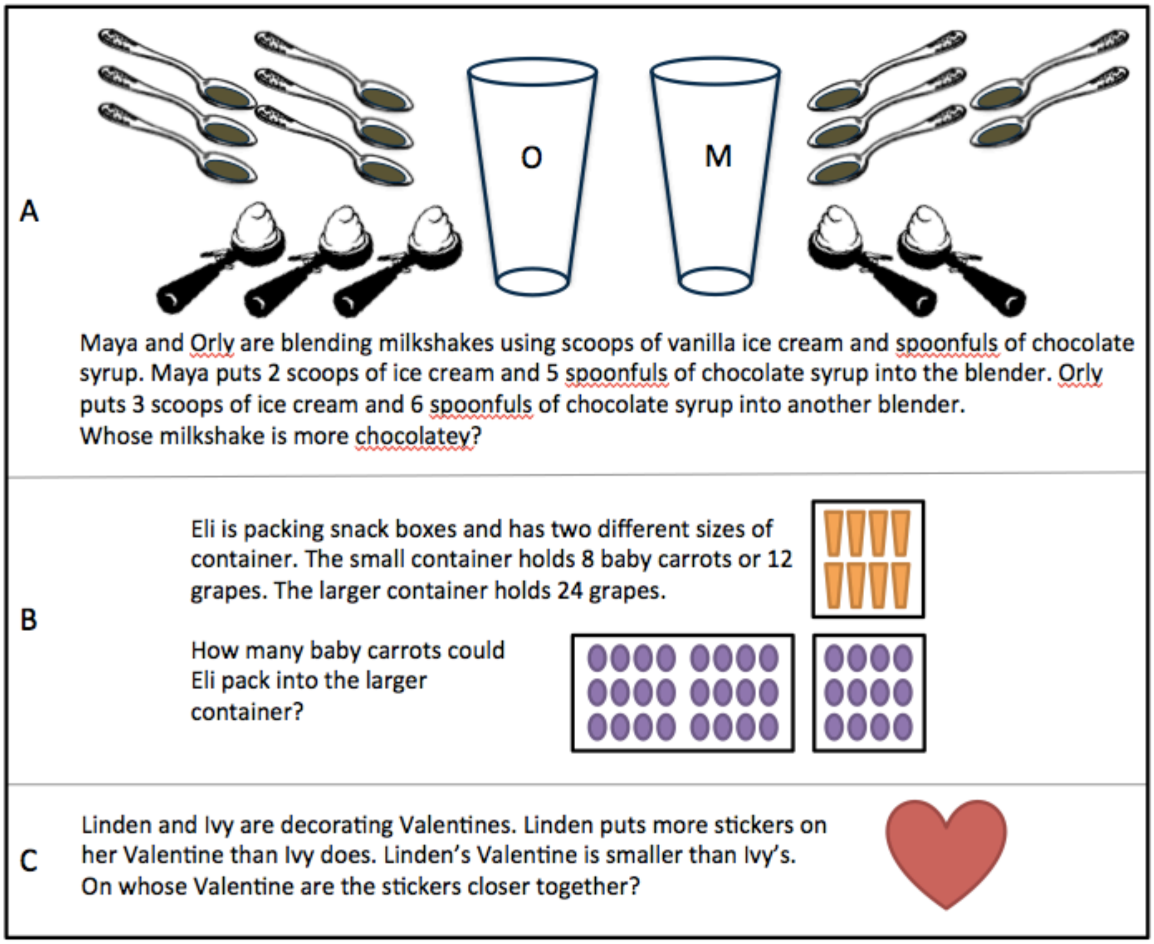


Fig. 1. Examples of proportional reasoning problems. A) numerical comparison, B) missing value, C) quantitative comparison.

The final approach is the cross-product strategy, which is an algorithm in which one denominator is multiplied by the opposite numerator and vice versa, and the products compared. This is a mathematically valid shortcut for creating equivalent fractions, but the procedure is often taught without that conceptual linkage and so students simply carry out the calculations without understanding their meaning.

English and Halford describe these problem types and approaches as evidence that proportional reasoning problems represent quaternary relations (1995). They make this point to distinguish proportional problems from classic verbal analogy problems of the form A:B::C:D, which they claim do not represent quaternary relations. Their claim is that classical analogies are solved by reducing the binary relation between the items in each pair to its essential relationship, and then comparing those two reduced relationships (although, as discussed below, classical analogies may retain their quaternary representation in some cases). Using relational reasoning terminology, the binary relation defining each pair of items is a first-order relationship, and the comparison of those two first-order relationships is a second-order relationship. The term second-order only defines the content of the relation, in that the items being compared in a second-order relation are themselves relations, and does not increase the complexity of the comparison. Thus, one way to solve a classic verbal analogy involves a series of binary comparisons.

However, the solution approaches for proportional reasoning problems also may represent a series of binary comparisons, and not necessarily a quaternary relation. The unit rate strategy allows one to calculate a single rate that defines the relationship between the numerator and denominator in one proportion, and compare that to the rate given in the other proportion. This solution strategy is congruent to the steps involved in solving a classic verbal analogy: first, one must reduce a numerical ratio to a single-value rate, repeat for the second ratio, and then compare two rates. Similarly, in the factor-of-change strategy, a learner reduces the multiplicative relationship between the two proportions' denominators to a single factor, and applies or compares that factor to the relationship between the given numerator(s). Only the quantitative comparison problems, in which there are no numerical values and therefore no reducible factors or rates, is necessarily a quaternary relation. Whether a learner solves the other problem types using a series of binary comparisons or a quaternary comparison depends on the affordances of the numbers given in the problem (whether they have common factors) and the skill of the learner (whether she or he has the cognitive flexibility to determine and carry out the optimal approach).

Analogical and relational reasoning theories

Therefore, if we think of proportional reasoning problems as being solvable either as a series of binary comparisons or as quaternary comparisons, they then become equivalent to classic verbal analogies, and can be characterized and analyzed using the existing wealth of knowledge about classical analogies. Contrary to the difficulty posed by learning formal proportional reasoning, analogical reasoning is used spontaneously by children to learn about their physical and natural environments (Alexander, White, & Daugherty, 1997; Inagaki & Hatano, 1987). Investigating proportional reasoning through the lens of analogical reasoning may provide insights into the cognitive challenges of proportional reasoning and potential ways to ameliorate them within pedagogical practices. To briefly summarize the analogical reasoning framework, analogies are generally characterized as a second-order relationship between two first-order pairs, and there are several prominent theories as to how those relations are mapped onto each other and evaluated.

Gentner's structure-mapping theory posits that individual components within each analog are aligned to each other, and then the structure of the base analog is used to make inferences about the target analog (Gentner, 1983). In this alignment process, the primary focus of attention is on the correspondence of components in the two structures; thus, the emphasis in comparison is between analogs. In a classic verbal analogy, the cross-analog relationships are not necessarily reduced to a single dimension, and all components maintain separate representational identities; therefore, this solution strategy, contrary to Halford's claim, may represent a quaternary comparison. However, when applied to a proportional reasoning context, the cross-proportion numerical relationships may be more easily reducible, depending on the numbers involved. If reducible, this solution strategy is congruent to the factor-of-change strategy described by Cramer, Post, and Currier, in that the primary focus of attention is on the relationship between aligned components in different proportions. As such, the multiplicative factor between denominators is applied to the multiplicative relationship between numerators (or vice versa), and evaluated for equivalence or comparison.

An alternate framework for analogical reasoning is called project-first (Sternberg, 1977), wherein the relationship between the items in the base analog is first identified and abstracted away from the individual entities. Then the learner creates a mapping between the first terms in each pair, and finally projects the relationship from the first analog onto the items in the target analog, without necessarily aligning all components across structures, or identifying specific relationships between corresponding components. Therefore, the primary focus of attention is on the binary relations within pairs.

The project-first solution strategy is similar to Cramer, Post, and Currier's description of the unit-rate strategy for proportional reasoning problems, according to which the rate is calculated within a given proportion and then projected onto the target proportion for completion or evaluation. Although the correspondence between items across proportions is not the focus of this strategy, it is of course necessary to correctly identify which of the given numbers is the numerator and denominator in each proportion, just as it is essential to identify the direction of a relationship in a verbal analogy. The relationship between "cat" and "mouse" is quite different to the relationship between "fly" and "spider" – the former describes "predator" or "eats" while the latter describes "prey" or "eaten by." Thus the structure is important, but the exact relationship between "mouse" and "fly," or the quantification of the multiplicative factor between numerators, is not preserved.

Prior analogical research has shown that adults, who have more experience or knowledge in a domain than children or novices, are more apt to use a project-first than a structure-mapping approach to solve classic semantic analogies (Thibaut, French, Missault, Gerard, & Glady, 2011). Neither of these theories have yet been fully explored in the context of proportional reasoning. Rather, prior research using a fraction comparison task has been characterized only in mathematical terms. In so doing, the prior results support a hybrid model of numerical processing, in which fractions are represented both as individual components and as holistic magnitudes and either representation can be accessed in response to task demands (e.g., Meert, Grégoire, & Noël, 2009, 2010a; Obersteiner & Tumpek, 2016). However, the fraction comparison task could be equivalently described as a proportional reasoning task, and used to investigate how people use within-fraction estimation and/or between-fraction comparisons. Therefore, the fraction comparison task may be a fruitful setting in which to apply these theories of analogical reasoning directly to a proportional reasoning context, thereby clarifying the cognitive challenges associated with proportional reasoning, and revealing the most successful approaches to those challenges.

Development of Proportional Reasoning

There are developmental theories associated with each of the perspectives discussed above – relational complexity theory and analogical reasoning theory – as well as educational theories about the development of mathematical knowledge. Due to the different philosophies and historical foundations of the research in cognitive psychology versus education, as well as the vastly different methods currently used to evaluate these theories, it is difficult to draw direct parallels between them. However, they are both empirically supported and their disparate lenses provide valuable insights into different aspects of mathematical learning. I will first address the cognitive theories previously presented, applying their developmental theories to proportional reasoning. Then I will present a very brief overview of the mathematics education perspective on proportional reasoning, to demonstrate the value added by jointly considering perspectives from the two disciplines.

Relational complexity theory

English and Halford claim that relational complexity is the primary source of variability in children's performance across a wide range of ages and tasks, above other effects such as knowledge and presentation style (1995). As such, they recommend taking relational complexity into account when designing all instructional activities. With respect to proportional reasoning, English and Halford state that children are first able to handle direct covariation between two quantities, such that no more than binary comparisons should be presented to children in lower elementary grades. That said, proportions themselves are binary relations, and due to children's observed difficulty with the notation, they suggest a potential advantage in introducing them earlier than in typical curricula, independent of the fraction notation and only in a conceptual context. Doing so would allow children to gain practice in attending to the quantitative relationships between parts and wholes, or to the relative magnitude represented by the proportion, without being hindered by numbers that could invoke the whole number bias (Ni & Zhou, 2005), or by fraction notation.

A similar line of research investigating continuous versus discretized representations of proportions, also without formal notation or symbolic numbers, shows that even young children are able to match proportions when represented continuously instead of with demarcated units (T. W. Boyer, Levine, & Huttenlocher, 2008). The authors' interpretation is that the units invoke numerical instead of proportional thinking. An alternate explanation could be that the continuous representations highlight the relative magnitude of proportions instead of the binary numerical relationship, thereby decreasing the complexity of the proportion from two quantities to one visually-perceived magnitude. In this way, younger children are able to handle the reduced complexity of non-numerical proportional comparisons, since, although technically quaternary relations, they are perceived as binary comparisons.

English and Halford also note that the 'effective complexity' of a problem may be different from the represented complexity, if only a subset of the given information needs to be taken into account. For example, a proportional comparison between 1/3 and 2/3 only requires attention to the numerators, not the denominators. Therefore, although the problem structure represents a quaternary relation, the effective complexity is only a binary relation. This qualification to the theory suggests the opportunity to introduce similar lower-complexity problems to children at younger ages as well, and reserve full quaternary relations for more advanced learners.

Analogical reasoning theory

Three phenomena borne of analogical reasoning research may also be applicable to proportional reasoning. First, a 'relational shift' has been documented in children's judgments, meaning that younger children make judgments based on perceptual cues while older children base their judgments on relations between items (Gentner, 1998). For example, given a question like that shown in Figure 2, younger children are more likely to select the perceptual match than the relational match. This is a distinct phenomenon from the relational complexity theory advanced by English and Halford, and yet it aligns with their suggestion that younger children are not capable of attending to relational information due to the processing load involved in binary comparisons.

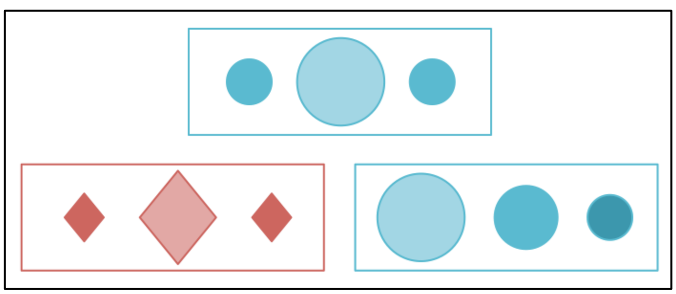


Fig 2. Example relational match-to-sample task. Standard given at top and options on bottom. Relational match on left, perceptual match on right.

However, subsequent research in analogical reasoning demonstrated that young children do in fact have the ability to make relational comparisons, provided they have sufficient conceptual knowledge to extract the pertinent relationships (Goswami, 1992). Tests such as the Test of Analogical Reasoning in Children (TARC) and the Mathematics Reasoning Test for Young Students (MRTYS), designed explicitly for early elementary students, supported the conceptual knowledge theory because it showed children's success in making relational judgments over perceptually similar distracters (Alexander & Buehl, 2004). The tests designed for young children tend to use qualitative size differences or numbers less than five, which are easily accessible to young learners. Most traditional proportional reasoning tasks used in educational settings involve larger numbers, and thus the conceptual knowledge theory suggests that quantitative problems should be deferred until children have a solid conceptual understanding of larger quantities.

A final analogical reasoning phenomenon has not, to my knowledge, been tested in a mathematical domain, but may also be applicable to proportional reasoning. Rattermann and Gentner (1998) demonstrated that relational labeling, such as daddy/mommy/baby to indicate patterns of size within sets of objects, promoted the use of relational features over perceptual matches that did not fit the size pattern. This approach successfully promotes young children's spatial reasoning (Loewenstein & Gentner, 2005), and may also be useful in quantitative reasoning. Fractional quantities have their own built-in labels, such as halves, thirds, and quarters, that describe the relation of numerator to denominator. It is worth testing the conjecture that using these labels with respect to different representations – including sets, lines, and areas, but excluding fractional and symbolic notation – promotes children's attention to relative quantities. Emphasizing fractions' relational 'names' may also be helpful in learning fraction operations, since operations such as addition and subtraction can only be carried out when the fractions have like denominators.

In the field of cognitive neuroscience, the study of relational reasoning encompasses analogical, categorical, and transitive reasoning, and a common brain region that emerges from fMRI studies is the rostrolateral prefrontal cortex (RLPFC). The RLPFC has been shown to be preferentially active on tasks that require second-order comparisons, when contrasted with those tasks that require only a series of first-order comparisons (Wendelken, Chung, & Bunge, 2011). A developmental comparison of children to adults demonstrated that the RLPFC exhibits less differentiation between second-order and first-order tasks in children than in adults (Crone et al., 2009). Similarly, a comparison of younger to older children and to adults showed that the RLPFC exhibits increasing selectivity for second-order tasks across age groups. These findings support and extend those discussed above: they show that children are capable of making higher-order relational judgments, but that they do so with less neural efficiency than adults. Although these neuroscientific conjectures are yet to be tested in the mathematical domain, applying these findings to proportional reasoning corroborates the suggestions made above, in that young learners need cognitive as well as quantitative supports in approaching a mathematical reasoning task.

Recent research by Richland and colleagues embarks on the direct application of these analogical frameworks to mathematical settings, showing that direct comparisons, gestures that highlight relevant relationships, reduction of working memory demands, and other cognitive supports do indeed facilitate learners' mathematical reasoning (Richland, Begolli, Simms, Frausel, & Lyons, 2016). As discussed below, this analogical perspective can be applied to facilitate learning in the domain of proportional reasoning.

Modern education theory

Switching away from cognitive psychology research, modern education research tends to focus less on the cognitive demands inherent to a particular topic of learning, and more on the social and environmental factors that promote or inhibit that learning. Much of the current educational research is rooted in constructivism, the theory that knowledge must be constructed by the individual by assimilating new information into existing mental constructs, and making accommodations to mental constructs when new information is incommensurate with them (Piaget, 1970). Constructivism contrasts with pre-Piagetian notions that knowledge is acquired only by associating stimuli to specific responses, as in behaviorism and associationism, and reflects both psychological and educational advances. Additionally, in the decades since Piaget, Vygotsky's theories have garnered attention; a few particularly influential ideas are that 1) learning is inherently social in that development occurs via the interaction between a child and his social environment, and 2) that the study of development should focus on the process of learning, not the outcome of knowledge attained (Vygotsky, 1978). Therefore, current education research can often be considered both neo-Piagetian and neo-Vygotskian, and typically has several levels of analysis, including not only individual construction of knowledge but also how that interacts with features of the social world such as the depth of mathematical discourse within a classroom, the power dynamics within dyads during partner work, or the teacher's responses to students' contributions, to give just a few examples.

As a result of this intense focus on <u>what</u> knowledge is constructed and <u>how</u> it is constructed, education research tends to be rooted within particular disciplines to a greater extent than is cognitive psychology research. Baroody and colleagues noted that "[developmental] hierarchies based on information-processing theory often do not adequately take into account or describe conceptual development, particularly the micro-conceptual development that is the bread and butter of classroom instruction" (Baroody, Cibulskis, Lai, & Li, 2004, p. 234). Thus, while the relational reasoning theories discussed above can be applied to multiple domains and contexts, the educational theories referenced in Section 2 are specific to the development of rational number knowledge.

The research I describe in Section 2 reflects the complex foundations of education research. The curriculum being examined, called Learning Mathematics through Representations (LMR), was designed as a progressive coordination of key mathematical concepts, ordered in such a way as to promote students' construction of knowledge about rational numbers (Saxe, Diakow, & Gearhart, 2012). The methods included individual work but also relied heavily on social learning in the form of classroom discussion, partner work, and community agreements on normative mathematical definitions. The LMR curriculum was developed as the culmination of a series of tutorial studies, teacher consultations, and incremental interventions, and emphasized the use of the number line in representing both integers and fractions. Many prior researchers have bemoaned the disconnect or even interference between students' integer and fractions knowledge (English & Halford, 1995; Mack, 1993; Ni & Zhou, 2005) so one of the goals of LMR, in using the number line to represent both integers and fractions, was to foster a connected understanding of rational number. The LMR curriculum was shown to be successful in this goal, with a large-scale efficacy study (over 500 students in 21 classrooms randomly assigned to LMR or district standard curriculums) showing substantial gains in rational number knowledge over the course of the LMR curriculum, that persisted through the end of the school year when LMR students' scores remained higher than comparison students' (Saxe et al., 2012).

The key concepts in LMR related to the number line were 'order' (numbers increase in value to the right and decrease in value to the left) and 'unit interval' (each unit is represented with an equivalent distance on the line). In addition to those representational concepts, the LMR curriculum advanced a number of additional key concepts that must be coordinated in order to understand a fraction: dividing a unit into equal intervals creates subunits; these subunits comprise the denominator of a fraction; the number of subunits in a measured distance comprises the numerator of a fraction; and equivalent fractions are in the same place on the number line but have different subunits. Proportional reasoning involves comparing two fractions, and thus requires even more conceptual coordination. Therefore, from a modern education research perspective, proportional reasoning is just as developmentally complicated as it is described in the cognitive psychology literature, if not more so.

The LMR research team investigated the ways in which the above key concepts and their mathematical definitions were generated over the duration of the curriculum. One of several analyses of the social construction of these ideas focused on the interplay between the normative mathematical definitions – the target knowledge – and informal gestures that helped the class reason about these mathematical ideas (Saxe, De Kirby, Kang, Le, & Schneider, 2015). The selected gestures were not specific to this mathematical context and yet fit the concepts well, so the teacher encouraged their use. In particular, the teacher both demonstrated certain gestures when introducing a new mathematical concept, and as the concept became part of the classroom community's shared knowledge, the associated gesture was invoked when the definition was referenced in the course of reasoning about a problem. Thus, the interaction between formal and informal ways of knowing supported the classroom community's shared understanding of these new concepts. I highlight this analysis as an example of how modern education research is both neo-Piagetian and neo-Vygotskian, and utilizes multiple levels of analysis.

As is apparent from the differences in language, foci, and level of analysis between cognitive psychology and education research, these perspectives are difficult to combine into a comprehensive model or theory. However, each of the theories represented above has been empirically supported, albeit in different contexts, and contributes valuable explanations regarding the development of proportional reasoning. The current state of general deficiency in proportional reasoning, combined with the scale of the problem and the intractable natures of both education policy and child-rearing culture, mandates that reformers press forward with all available potential solutions. To do so, we must draw not just on one successful theory or another, but on all ideas for potential improvement.

Cognitive and Social Analyses of Proportional Development

This dissertation reflects the idea that both cognitive and social angles on learning theory are relevant and necessary for educational improvement. There are important findings from cognitive psychology that may illuminate instructional design, and there are important interactive aspects of educational environments that affect how that instruction is interpreted by teachers and students. Real learning is the coordination of knowledge within mental constructs that occurs in social, noisy classroom environments. Neither level can be ignored, although the nature of the questions regarding each dictates that they be addressed separately. Section 1 of this dissertation contains reports of two experiments in the cognitive psychology tradition that investigate the strategies used by novice and proficient mathematicians as they solve proportional reasoning problems. These experiments use eye-tracking technology to record the location of participants' gazes and how their eyes move between the numbers presented on the screen. This experiment and methodology allowed me to ask whether children approach these problems differently than adults do, and whether those differences are due to mathematical proficiency or cognitive maturation. The findings are reported in Chapter 1. In Chapter 2, I ask whether adults' strategy choice and flexibility, specifically regarding strategies that align with those outlined in the analogical reasoning literature, lead to performance benefits on a proportional reasoning task. Section 2 contains research in the modern education tradition, although the analysis described takes a quantitative approach that is not typical of that tradition. In analyzing students' mistakes on assessment questions related to fractions, I investigate how their coordination of ideas is impacted by differences in curricula and classroom environments.

Section One Introduction

An Introduction to Gaze Analyses

For over two centuries, scientists have probed the workings of the mind by measuring movements of the eye (Wade, 2015), but the modern study of eye-tracking as indication of cognitive processes was launched by Yarbus in 1967. He provided a simple illustration that gaze reflects cognition by asking subjects different questions as they viewed the same painting. When asked to judge the age of each character, a sample participant looked primarily at the depicted faces; when asked to judge the material wealth of the family, he looked primarily at the characters' clothing and some of the surrounding objects. The differences in recorded eye movements, or saccades, support the intuition that the eyes are directed towards the object of one's thoughts.

Although it is possible to attend covertly to a spatial location without moving one's eyes to it, it is more common – and more effective – to fixate that which we attend to. For example, Deubel and Schneider (1996) cued a saccade to a specific location and measured visual discrimination at a point either coincident or adjacent to the saccade target. When the cued saccade location coincided precisely with the discrimination target, accuracy was very good. Otherwise, even when the participants knew in advance the location of the upcoming discrimination target, they performed at chance. Thus, they were unable to dissociate covert visual attention from saccade location, demonstrating an inextricable link between eye movements and cognition (Deubel & Schneider, 1996). Further corroborating the close link between attention and gaze, fMRI investigations have found that the frontal eye fields (FEF), which control eye movements, are also implicated in the deployment of covert visual attention (Corbetta et al., 1998; Grosbras, Laird, & Paus, 2005). Direct cortical stimulation in macaques also supports the neurobiological link between visual attention and gaze. Researchers stimulated macaques' FEF (Moore & Fallah, 2004) and superior colliculus (Müller, Philiastides, & Newsome, 2005), two regions known to play a part in oculomotor planning, with an electrical current that was not strong enough to elicit a saccade. Both studies found that the macaques were better able to discriminate or detect stimuli presented at the preferred location of the neurons being microstimulated, despite the eyes remaining still. Based on the evidence presented in their review, Awh, Armstrong, and Moore (2006) concluded that oculomotor saccade selection and visual attention are guided by the same neural substrates, FEF and SC.

An interesting implication of this finding is the ability to manipulate visual attention in order to alter cognition. The most famous study of this phenomena is Grant & Spivey's analysis of people solving (or not solving) the classic Duncker's radiation problem and analogs of it (2003). They identified the area of an associated diagram that successful participants spent the greatest proportion of a trial studying, and then made that area more visually salient by making the critical object pulse in thickness on the screen. In the follow-up experiment using the animated diagram, a significantly greater proportion of the participants solved the insight problem without verbal hints than in the original experiment with the static diagrams. Thus, the external manipulation of visual attention increased people's insight on a difficult problem-solving task.

Gaze Analysis as a Measure of Cognitive Processes

Given the strong link between eye gaze and attention, quantifiable measures of gaze, such as duration and location, can be useful indicators of underlying cognitive processes. An ample body of research has shown that people make more frequent saccades towards, and fixate longer on, the components of a visual stimulus that require their attention. These may be the components that are most difficult to decode, or those that are most visually dense, or those that are most informative for the task at hand.

In the study of silent reading, for example, very short words are often not fixated at all, while longer words almost always are, and people often go back and refixate on words that have more letters or are more difficult to comprehend (Rayner, 1998). Eye gaze metrics reflect both task differences, such as misleading sentences or differences in character sets between languages, and individual differences in skill level. For example, beginning readers or less-skilled readers make longer fixations, shorter saccades, and more refixations than skilled readers (Rayner, 2009). These examples illustrate the validity of eye gaze metrics as proxy measures of cognition, because the words or stimuli that are fixated more frequently or longer are those that require more cognitive processing.

Beyond providing additional quantification of known differences between tasks and readers, gaze analysis can inform models of cognitive processes. In an experiment involving linguistic predictions and inferences, Calvo (2001) found that working memory performance on an independent task modulated subjects' refixations while reading ambiguous sentences. Specifically, subjects with high working memory capacity exhibited fewer refixations to the prior sentence when processing a subsequent inference, than subjects with low working memory capacity, indicating that domain-general working memory is leveraged to facilitate linguistic processing. Similarly, Traxler, Williams, Blozis, and Morris (2005) found that working memory capacity impacted the time taken in refixations while reading syntactically challenging sentences. These two findings illustrate how gaze measures not only test but also inform theoretical cognitive models.

Gaze analysis has been used extensively in reading research, but has also provided insights regarding other cognitive skills. For example, in a study of mental rotation, response times tend to increase linearly as a function of angular disparity (Just & Carpenter, 1976). Just and Carpenter found that the number of saccades between two matching figures also increased linearly as a function of angular disparity, but not uniformly throughout all stages of the trial. The increase in angular disparity primarily impacted the number of saccades during the 'transformation and comparison' stage of the trial. The other stages – initial visual search and subsequent confirmation – were associated with only a slight increase in number of saccades as angular disparity increased. Therefore, gaze analysis made it possible to pinpoint the specific cognitive mechanism impacted by the task manipulation – something that could not have been achieved on the basis of response times alone.

In summary, the spatial and temporal data in gaze measures provide additional information beyond a task's behavioral measures to illustrate the cognitive processes involved in a task. The two experiments reported in this section use counts of different types of saccades to characterize the problem-solving strategies participants use when selecting one of two fractions with the larger magnitude.

Chapter 1 – Using Eye Tracking to Probe Developmental and Skill-based Differences in Fraction Magnitude Evaluation

Co-authored with undergraduate Honor's Student Jeffrey L. Crawford

Introduction

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, United States algebra teachers 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). More generally, recent reports of difficulty with fraction comprehension (Kilpatrick, Swafford, & Findell, 2001; Mazzocco & Devlin, 2008; Stafylidou & Vosniadou, 2004) are cause for concern. 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. The introduction to fractions in a school curriculum represents students' first exposure to rational numbers, and the first time since they first learned to count that the rules of counting numbers are violated. Rather than adhering to a one-to-one mapping between numeric symbols and magnitudes, an infinite number of fractions can represent a magnitude – and there is an infinite number of interstitial magnitudes between any two given fractions (Vamvakoussi & Vosniadou, 2004). Additionally, fractions can represent many types of numerical relationships. The notation 3/5 can mean three pieces out of one object that's been divided into five sections, or it can mean three objects out of a set comprised of five objects. These are distinct meanings, but both are examples of the simplest definition of fractions, which is that they represent parts of a whole. Fractions also represent the quotient of a division, a proportional relationship represented as a ratio, a relationship between two quantities that defines a rate, and an operator that can be applied to other numbers, e.g., the function in which x is 2/3 of y (Behr et al., 1983). Furthermore, all of these definitions are encompassed in the same a/b notation, obscuring both the intended meaning and the holistic magnitude.

Many elementary school curricula focus on just one or two of the above definitions of fractions, making the material easier to learn, but leaving many students with an incomplete understanding. As a result, students learn to carry out procedural algorithms that allow them to correctly answer many problems involving fractions. However, when faced with a problem that tests their conceptual understanding, or exceeds their procedural skill level, such as estimating the sum of 11/12 and 7/8, the majority of high school students select an incorrect answer (Carpenter, Corbitt, Kepner, Lindquist, & Reys, 1980).

Fraction comparison task as an indicator of a mental representation of fractions

One of the core components of fraction proficiency is a working knowledge of fraction magnitudes (Bright, Behr, Post, & Wachsmuth, 1988; Kilpatrick et al., 2001; National Mathematics Advisory Panel, 2008), which can be tested by asking students to indicate which of two fractions is larger. This fraction comparison test has been used to investigate fraction understanding in adults (Faulkenberry & Pierce, 2011; Ischebeck, Schocke, & Delazer, 2009; Meert et al., 2009, 2010a; Sprute & Temple, 2011) and children (Behr, Wachsmuth, Post, & Lesh, 1984; Clarke & Roche, 2009; Meert, Grégoire, & Noël, 2010b; Smith III, 1995; Vukovic et al., 2014). A variety of strategies can be successfully applied to compare two fractions, ranging from the simple strategy of selecting the larger numerator when the denominators are identical, to more computationally complex strategies such as converting both fractions to decimals.

More generally, two distinct mental representations have been thought to influence strategy choice on the fraction comparison task. The first, that fractions are represented as an integrated magnitude, is commonly referred to as the magnitudes or holistic processing theory. This type of processing would be adaptive when comparing highly familiar fractions (e.g., 3/4 and 2/5), because it is possible to draw on prior knowledge rather than specifically comparing the individual components of the fractions. By contrast, the second proposed mental representation is that the numerator and denominator of each fraction are represented as their separate components, and is referred to as the componential theory. For example, a comparison between 5/6 and 9/12 might prompt one to notice that 12 is two times 6, but that 9 is less than twice 5. Prior research has provided empirical support for both the componential theory (Bonato, Fabbri, Umilta, & Zorzi, 2007; Ischebeck et al., 2009) and the holistic theory (Faulkenberry & Pierce, 2011; Schneider & Siegler, 2010; Sprute & Temple, 2011) of fraction representation in adults.

Additional studies have illustrated behavior compatible with both componential and holistic processing depending on the task context, suggesting a hybrid mental representation in adults that allows for flexibility in strategy selection (Faulkenberry, Montgomery, & Tennes, 2015; Fazio, DeWolf, & Siegler, 2015; Gabriel et al., 2013; Meert et al., 2009, 2010a). For example, in Meert, Grégoire, and Noël (2009), participants were presented with pairs with the same denominator (e.g., a/x vs b/x) as well as pairs that shared a numerator (e.g., x/a vs x/b). Their response times on same-denominator pairs were compatible with the componential theory, whereas their response times on the same-numerator pairs were compatible with the holistic theory. These results highlight adults' ability to switch processing modes, and thereby access different mental representations, depending on the task at hand. Indeed, the ability to flexibly shift between strategies is indicative of a participant's proficiency and success in solving other mathematical tasks (Fazio, DeWolf, & Siegler, 2016; Smith III, 1995).

Eye-tracking technology as a means of characterizing strategy

The aforementioned studies have used highly precise behavioral and chronometric methods to make inferences about mental representations, 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).

The first published studies to use eye-tracking with a fraction comparison task illustrated how participants adapted their gaze behavior to different pair types (Huber, Moeller, & Nuerk, 2014; Obersteiner et al., 2014). Specifically, Obersteiner and colleagues demonstrated that participants fixated longer on numerators when the denominators of the two fractions being compared were the same, and vice versa. Similarly, Huber, Moeller, and Nuerk (2014) analyzed first fixations during blocks of same-type trials and noted that when the numerators were always the same across pairs, participants looked first at the denominators to encode the relevant information, and vice versa for blocks of same-denominator trials.

Further support for the hybrid model comes from analyzing patterns of saccades between numerals in addition to eye fixations on specific numerals. Obersteiner and Tumpek (2016) and Ischebeck, Weilharter, and Körner (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 additionally found that saccades between the numerator and denominator within the same fraction were more common when the fractions shared no common components (2016). These initial eye-tracking findings lend support to the hybrid representation theory, as they show that adults use componential strategies when they are adaptive, and holistic strategies when all digits need to be taken into account.

Therefore, both behavioral and recent eye-tracking studies support the hybrid model of processing in adult participants. The fraction comparison studies involving children have used interview techniques to elaborate various strategies used (e.g., Clarke & Roche, 2009; Smith III, 1995); to our knowledge, none have addressed the question of mental representation of fractions, nor probed strategic approaches using eye-tracking. In the current study, our aim is to compare strategy use between children and adults by analyzing their saccadic eye movements during performance of a fraction comparison task. Critically, we seek to disambiguate age-related differences related to maturation and experience. Not only do adults have more exposure to numbers and fractions than children, both in and out of school, they also have advantages over children in many general cognitive skills. For example, adults have faster processing speed (Kail, 2000), better working memory, more cognitive flexibility, and higher levels of inhibition and sustained attention (e.g., Bunge & Wright, 2007; Case, Kurland, & Goldberg, 1982; Geier & Luna, 2009; Luna, Garver, Urban, Lazar, & Sweeney, 2004) – all of which are likely to impact their fraction comparison ability. Thus, the tandem influences of maturity and expertise can make it difficult to interpret whether developmental findings reflect changes due to age or skill acquisition.

In addition to these age-related differences, it is also possible that the mathematical strategies used by children as they are learning new concepts are qualitatively different than those used by experienced adults. For example, a set of studies using the number line magnitude placement task documented the use of increasingly mature 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, showing that task proficiency corresponded to increasingly mature strategies. For the fraction comparison task, proficient adults' strategies have been documented (Ischebeck et al., 2016; Obersteiner & Tumpek, 2016) but it is not yet known how cognitive maturity interacts with these strategies, or if children adopt different strategies entirely.

Eye movements are influenced by both maturity and expertise

Beyond fraction tests, eye-tracking studies involving other cognitive tasks have documented effects of both cognitive maturity and proficiency on saccadic eye movements. Age differences in eye movements are well founded for a variety of task-based approaches (Bucci & Seassau, 2012; Cohen & Ross, 1977; Fukushima, Hatta, & Fukushima, 2000; Irving, Steinbach, Lillakas, Babu, & Hutchings, 2006; Irving, Tajik-Parvinchi, Lillakas, González, & Steinbach, 2008; Munoz, Broughton, Goldring, & Armstrong, 1998; Ross, Radant, Young, & Hommer, 1994; Salman et al., 2006). The most consistent finding related to general cognitive maturity is an age-related decrease in saccade latency, or the time elapsing between the appearance of a stimuli and the initiation of the orienting saccade (Bucci & Seassau, 2012; Cohen & Ross, 1977; Fukushima et al., 2000; Irving et al., 2006; Munoz et al., 1998; Ross et al., 1994; Salman et al., 2006). Similarly, the error rate for saccade inhibition during antisaccade tasks decreases with age (Bucci & Seassau, 2012; Fukushima et al., 2000; Irving et al., 2008; Luna et al., 2004; Luna, Velanova, & Geier, 2008; Munoz et al., 1998; Ross et al., 1994), which again indicates an effect of maturity in oculomotor control.

Skill level in various domains is also associated with characteristic eye movements. 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. As outlined above, the three prior eyetracking studies focused on numerical magnitude understanding (Schneider et al., 2008; Sullivan et al., 2011; van't Noordende et al., 2016) provide an example of the development of mathematical proficiency. Thus, the accumulation of knowledge and experience in working with fractions may also lead to strategic differences as well as general differences due to maturity.

The current study

To investigate the effects of maturity versus proficiency on saccades during a fraction comparison task, we divided our sample into three groups: high proficiency adults (HP adults), high proficiency children (HP children), and low proficiency children (LP children). First we examined behavioral performance on each of the task conditions for each group. We then tested for differences in saccade patterns between the groups, and measured the extent to which each group exhibited distinct patterns of eye movements between the conditions.

Based on the research described above showing a general improvement in oculomotor control 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 expected saccade patterns, i.e., the relative number of different types of saccades, to be related to level of proficiency. Thus we predicted that children who perform similarly to adults would demonstrate saccade patterns more similar to adults than to lower-performing children.

Additionally, based on prior research indicating that proficiency is associated with task-relevant behavior (Gegenfurtner et al., 2011), we predicted that the participants who demonstrate higher mathematical proficiency on the fraction comparison task would exhibit more task-relevant saccades than less proficient participants, and also greater strategic flexibility between task conditions. Thus we predicted that highly proficient groups would demonstrate greater differences in saccade patterns between task conditions, as well as more coherent gaze behavior within each condition.

Methods

Participants

We recruited 35 5th-grade children and 38 college students for this study. The children were recruited from a charter school in Oakland, California in a socioeconomically depressed community. 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 UC Berkeley who participated in the study for course credit in a Psychology course, as part of a larger study on adults' fraction strategies.

One child was excluded from the study on the basis of less than 50% valid eye gaze data. The final sample included 34 children (ages 9-11; mean = 10.6, sd = .53; 19 girls, 15 boys) and 38 young adults (18 to 22 years old, mean = 20.3, sd = 1.2; 9 male, 25 female, 4 declined to state). All participants had normal or corrected-to-normal vision.

Cognitive Measures

Stimulus and 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 one-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.

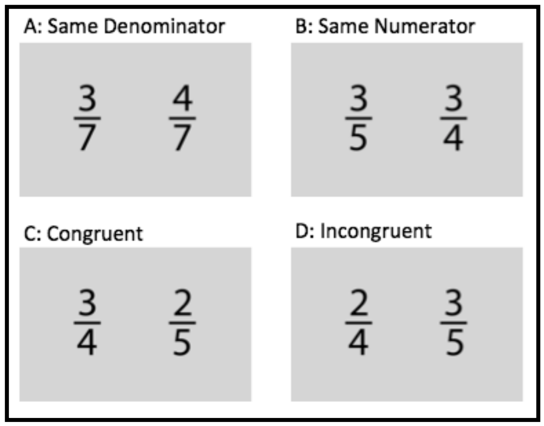
The experiment was conducted on a Tobii T120 eye-tracker, with a sampling rate of 120Hz (one measurement every 8.3ms). Participants were asked to sit in front of a Tobii eye-tracker at the recommended distance of approximately 64 cm. The session began with a 9-point calibration protocol to ensure that 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 x 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.

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 A). We used fraction pairs with a numerical distance of one between the denominators and between the numerators (e.g., 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).

There were 32 trials total, divided into 4 separate conditions with 8 fraction pairs each, adapted from Ischebeck, Schocke, and Delazer (2009). In these 8 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.

Fig. 1. Sample Items for each condition.



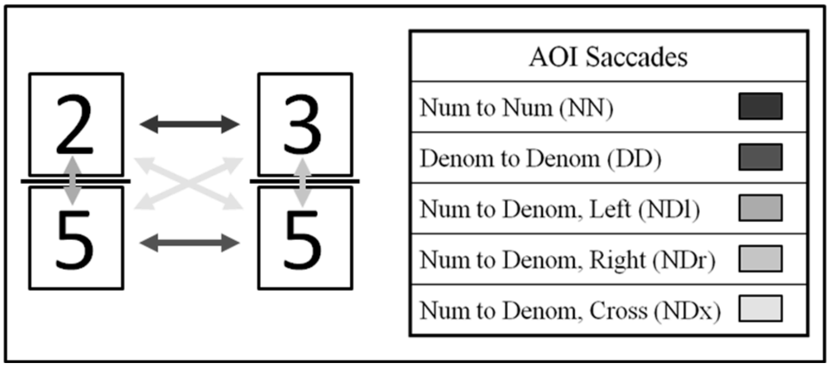
Following Ischebeck et al. (2009), we used four conditions that elicit distinct behavioral signatures. In the Same Denominator (SD) condition, fraction pairs had the same denominator but different numerators (Fig 1, panel A). This was the simplest condition, because when two fractions have the same denominator, then the larger fraction has the larger numerator. In the Same Numerator (SN) condition, each of the fraction pairs had different denominators, but the same numerators (Fig 1, panel B). 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 (Fig 1, panel C). 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 (Fig 1, panel D). Conditions were interspersed pseudo-randomly over the course of a single block of trials.

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 and eye gaze data suggest that participants considered both numerators and denominators on IC trials.

Metrics

From the Tobii output file we calculated trial accuracy and response times, as well as the number of saccades between digits per trial (saccades/trial). We defined an AOI for each digit on the screen, and measured saccades through the four areas of interest (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).

Fig. 2. Examples of AOI saccades between the four AOI: LN, LD, RN, RD, which are shown here as rectangles around the numbers.



Analytic approach

Preprocessing Steps

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). Any set of samples within a single AOI that lasted less than 40 milliseconds we assumed 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 100 milliseconds of missing samples were concatenated, under the assumption that the disruption to be caused by a blink. If data from both eyes were available, they were averaged to determine gaze location; however, a valid recording from one eye was sufficient for the Tobii software to determine which AOI the participant was fixating. All participants' data were screened to ensure that at least 50% of their data points were valid samples. As noted previously, this criterion led to the exclusion of one child participant.

For the gaze analyses, we normalized the participant-level differences in number of saccades by calculating z-scores for each saccade metric based on an individual's mean number of saccades across conditions. By this measure, 0 represents an individual's grand mean of all types of saccades across all conditions, 1 represents 1 standard deviation above the individual's mean, and -1 represents 1 standard deviation below the individual's mean. Only correctly performed trials were analyzed, to ensure that the analyzed saccade patterns reflected successful strategy use.

Grouping Criteria

To address our research questions about the differences due to age versus the differences due to proficiency, the adults and children were categorized into two groups according to their performance on the SD and SN conditions. These conditions require essential knowledge of the relationship between numerators and denominators. The high proficiency groups for both children and adults were defined as having at least 70% accuracy on both the SD and SN conditions. This set of criteria split the two original groups into a higher proficiency (HP) group of 33 adults, an HP group of 12 children, a lower proficiency (LP) group of 5 adults, and an LP group of 22 children. Given the small number of LP adults, we focus here on the 33 HP adults, 12 HP children, and 22 LP children. Comparisons across these groups allowed us to investigate age-related differences in eye saccade patterns between HP children and HP adults, as well as proficiency-related differences between HP and LP children.

Statistical Analyses

To test for general differences in performance and behavior, we used a 2 [group] x 4 [condition] ANOVA with repeated measures to test for differences in accuracy, response times (RTs), and average number of saccades per correct trial. We conducted these behavioral analyses separately for HP children vs. HP adults, to test for differences due to age, and for HP children vs. LP children, to test for differences due to proficiency. Next, to compare saccade patterns of HP children with those of HP adults as well as of LP children, we conducted a 2 [groups] x 4 [conditions] MANOVA using the z-scored values of the five types of AOI saccades as dependent variables. We selected this multivariate approach to test for characteristic differences in saccade patterns between conditions and groups, combining the five saccade types into one analysis. To further investigate differences in individual saccade types, we computed ANOVAs for each of the five saccade types in turn, testing for specific differences between groups and between conditions. Finally, to complement results from the MANOVA, we used a linear discriminant analysis to predict saccade patterns for each condition and determine classification accuracy for HP children and HP adults.

Because we included only correct trials in our analysis, sample sizes of each group differed across conditions. This is because some adults and children did not provide any correct answers for the 8 trials within a condition. Their data were omitted from analysis on those conditions where there was no viable data, but included in conditions where their data were present. Specifically, one person was omitted from the sample of HP adults on the IC condition, four HP children were omitted from the IC condition, and three, five, and four LP children were omitted from the SD, SN, and IC conditions respectively.

Results

Behavioral Results

Effects of Age

We defined our HP groups based on at least 70% accuracy in the two easiest conditions (SD and SN); nevertheless, we expected HP adults to demonstrate better accuracy than HP children on all four conditions. Indeed, we found significant differences in overall accuracy between HP adults and HP children, $F(1,43) = 26.32$, $p < .001$ in addition to differences by condition, $F(3,129) = 37.74$, $p < .001$. There was also a significant age group by condition interaction, $F(3,129) = 22.55$, $p < .001$, driven by a large drop in accuracy for HP children on the IC condition (Fig. 3). HP children achieved only 39% accuracy on the IC trials, even though they could have solved them correctly by identifying either the larger of the two numerators or the larger of the two denominators. Thus, the more advanced 5th-graders, who demonstrated proficiency in comparing numerators on SD trials or denominators on SN trials, attempted to integrate information about both the numerators and denominators to select the larger of two fractions, but had difficulty doing so. Adults also had greater difficulty with the IC condition than the other conditions (albeit not to the same degree), indicating that both HP children and adults attempted to integrate all four digits when selecting the larger fraction.

RTs also differed significantly by age group, $F(1,41) = 4.25$, $p = .045$, and condition, $F(3,124) = 12.58$, $p < .001$, although the interaction was not significant $F(3,124) = 1.11$, $p = .35$. HP children took longer to respond than HP adults, but displayed similar increases in RTs as the condition difficulty increased across SD, SN, CO, and IC trials (Fig. 4). Therefore, both HP groups were sensitive to condition difficulty, and the children were slower overall.

Effect of Performance

The HP and LP child groups were defined by a difference in accuracy on the SD and SN conditions. Nonetheless, we conducted an ANOVA including condition to better characterize group differences in accuracy profiles across all four conditions – that is, to test whether the LP children performed disproportionately worse on the mixed fractions (CO, IC) than HP children. This analysis revealed a main effect of condition, showing a decline in accuracy for both groups as condition difficulty increased, $F(3, 96) = 3.63$, $p = .016$, but also a main effect of proficiency group, $F(1,32) = 26.57$, $p < .001$, with HP children performing better overall. The interaction between group and condition was also significant, $F(3,96) = 4.91$, $p = .003$. Whereas the HP children scored highly on SD, SN, and CO conditions, and poorly on IC, the LP children did almost equally poorly on all conditions.

Despite differences in accuracy, HP and LP children showed minimal differences in RTs across the four conditions. There was no main effect of proficiency level $F(1,28) = .15$, $p = .70$, or condition, $F(3,80) = 0.45$, $p = .72$, and there was a non-significant trend towards an interaction $F(3,80) = 1.63$, $p = .19$. Therefore, in this age group, proficiency with the fraction concepts was not associated with greater efficiency; nor was there a speed-accuracy tradeoff between the groups (Table 1, Fig 4).

Table 1. F-values of the repeated measures ANOVAs between HP adults and HP children, and between HP children and LP children on accuracy, response times, and number of saccades/trial. * p < .05 ; ** p < .01; *** p < .001

	HP Adults and HP Children			HP Children and LP Children		
	Accuracy	RTs	Saccades	Accuracy	RTs	Saccades
Group	26.3 ***	4.2 *	11.9 **	26.6 ***	.14	2.7
Condition	37.7 ***	12.6 ***	5.1 **	3.6 *	.45	.71
Interaction	22.6 ***	1.1	.20	4.9 **	1.6	1.8

Fig. 3. Accuracy on the fraction comparison task conditions for HP adults, HP children, and LP children.

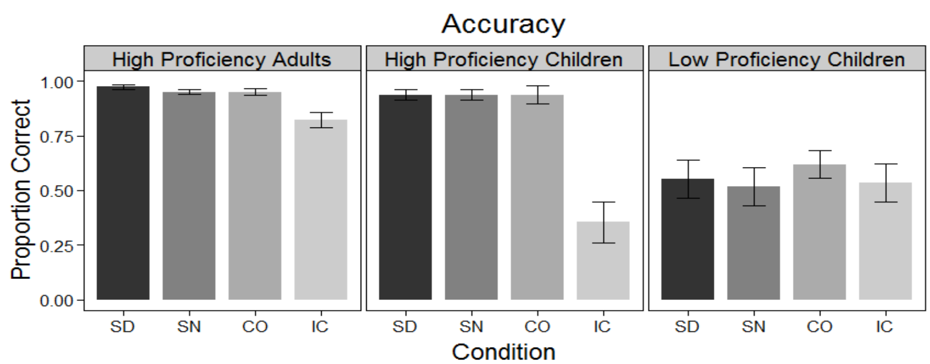


Fig. 4. Average response times (ms) on correct trials within each condition for the HP adults, HP children, and LP children.

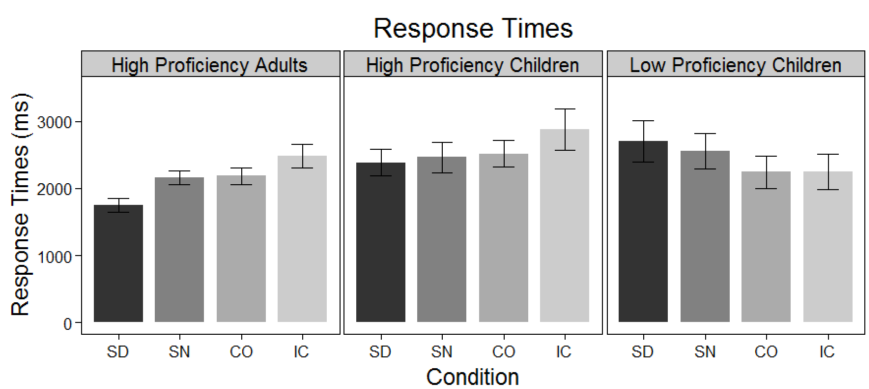
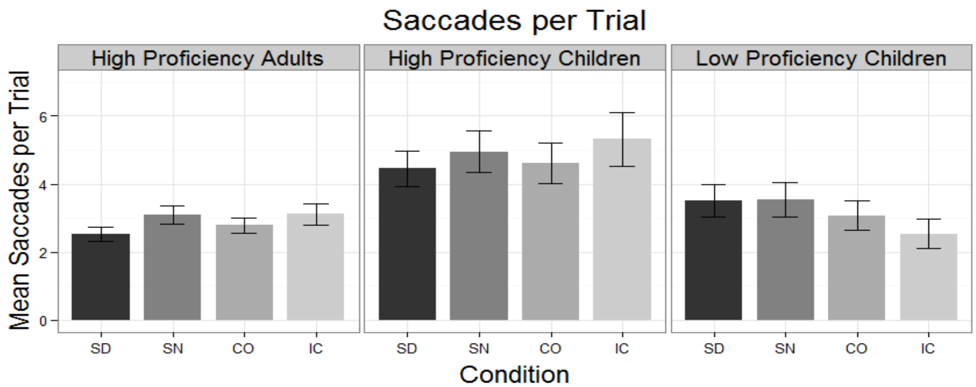


Fig. 5. Average number of saccades/trial on correct trials within each condition for HP adults, HP children, and LP children.



Gaze Analysis Results

Our first step in the gaze analysis was to test whether the number of saccades that individuals make while analyzing a stimulus array – i.e., saccadic efficiency – differs by proficiency and age. We conducted the same ANOVAs as used for accuracy and RT, with total number of saccades/trial as the outcome measure (Fig. 5). In the comparison of HP adults and HP children, saccades/trial mirrored the results for RTs. There were significant differences between these two groups, $F(1,41) = 11.93$, $p = .0013$, with adults making fewer overall saccades than children. There were also significant differences between conditions, $F(3,124) = 5.06$, $p < .001$, with both groups making more saccades on more difficult conditions. There was no interaction, $F(3,124) = .15$, $p = .94$. In the comparison of HP and LP children, saccades/trial did not differ by proficiency level, $F(1,28) = 2.75$, $p = .11$, or condition, $F(3,80) = .71$, $p = .55$, and there was a non-significant trend towards an interaction, $F(3,80) = 1.77$, $p= .16$. This outcome is similar to that for RTs, in that adults made fewer saccades than children, but there was no difference between the two groups of children. However, the HP children and adults both exhibited sensitivity to condition difficulty in that there were more saccades for the more difficult conditions.

Multivariate Analysis of Saccade Types

In order to determine whether there were differences in saccade patterns between our groups, we computed 2 [group] x 4 [condition: SD, SN, CO, IC] MANOVAs for effects of age and proficiency, with the five saccade types (NN, DD, NDl, NDr, and NDx) as dependent variables. For HP children vs. HP adults, there were differences in age group, $F(1,167) = 4.95$, $p < 0.001$, and condition, $F(3,167) = 2.35$, $p = 0.003$, without a significant interaction, $F(3,167) = 1.34$, $p = 0.17$. For HP children vs. LP children, there was a trend towards a difference in saccades/trial as a function of proficiency level, $F(1,112) = 2.25$, $p = 0.054$, as well as significant differences across conditions, $F(3,112) = 4.18$, $p < 0.001$, but no interaction, $F(3,112) = .41$, $p = 0.97$. Thus, all three primary factors – proficiency, age, and condition – affected the distribution of types of saccades in the different task conditions. To further characterize the differences indicated by the significant results in the MANOVAs, we conducted follow-up ANOVAs using each saccade type as an individual outcome measure.

Univariate Analysis of Saccade Types

Effects of Age. In the comparison of HP children to HP adults, the main effects of age group were driven by a greater number of NDl saccades for children, $F(1,41) = 4.52$, $p = .04$, and a greater number of NDx saccades for adults, $F(1,41) = 6.77$, $p = .013$. There was also an interaction between condition and group for NN, $F(3,124) = 5.61$, $p =.001$ (Figure 6). HP children made more NN saccades on SD trials than did HP adults. Noticeably, the interaction was characterized not only by the HP children showing a greater number of NN saccades on SD trials, but also fewer NN saccades on SN trials – i.e., greater differentiation across these conditions than adults. There was also a very significant main effect of condition for the DD saccades, $F(3,124)= 16.71$, $p < .001$, because both groups demonstrated a greater number of DD saccades on SN trials than on other conditions (Table 2, Figure 7). Taken together, HP adults and HP children showed different patterns of saccades across conditions, with adults making more NDx saccades overall and children making more NDl saccades overall. Both groups exhibited more attention to relevant digits on the SN condition, as expected, but only HP children showed preferential saccades on the SD condition.

Effects of Proficiency. The only identifiable differences between the HP and LP groups of children were between each of the conditions, not between proficiency groups. Increases in NN saccades for SD relative to the other conditions, $F(3,80) = 6.31$, $p <.001$, and in DD saccades for the SN condition, $F(3,80) = 9.75$, $p<.001$, were observed for both groups (Figure 7). The highly significant effect of condition, and the absence of a group difference, indicates that HP and LP children respond similarly to the stimuli.

Table 2. F-values of the repeated measures ANOVAs for the five AOI saccade types between HP adults/HP children and HP children/LP children. * <.05 , ** <.01, *** <.001

	HP Adults and HP Children				
	NN	DD	NDl	NDr	NDx
Age Group	0.03	1.0	4.5 *	0.40	6.77 *
Condition	1.7	16.7 ***	1.2	1.3	2.3
Interaction	5.6 **	2.3	0.9	0.78	1.3
	HP Children and LP Children				
	NN	DD	NDl	NDr	NDx
Proficiency Group	2.1	2.5	0.5	0.2	2.5
Condition	6.3***	9.7 ***	0.4	2.1	2.3
Interaction	0.8	0.2	0.6	0.3	0.3

Fig. 6. Saccades between the AOIs surrounding the numerators of the two fractions. The raw scores were mean-centered using a z-score transformation. Zero represents the individual's grand mean across all conditions.

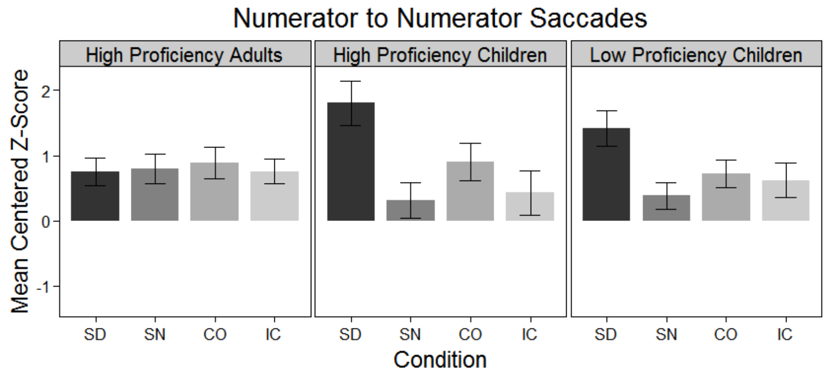
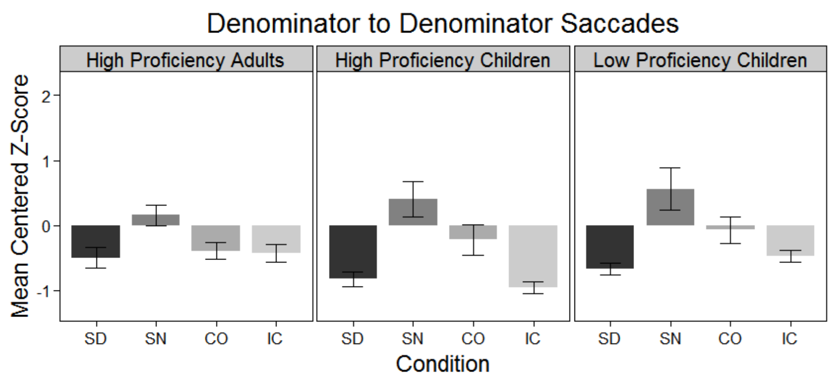


Fig. 7. Saccades between the AOIs surrounding the denominators of the two fractions. The raw scores were mean-centered using a z-score transformation. Zero represents the individual's grand mean across all conditions.



Linear Discriminant Analysis for Age Groups

We hypothesized that participants who demonstrated high proficiency in the fraction comparison task would exhibit distinct gaze patterns for each condition, because it has been previously established that flexible strategy use is associated with proficiency in many domains (Gegenfurtner et al., 2011). The MANOVA results showed that both HP children and adults did indeed exhibit differences in the frequency of certain saccades between conditions; we then performed a follow-up linear discriminant analysis (LDA) to determine which specific conditions elicited the most distinct gaze patterns. In this way, we used LDA to look at the overall pattern of eye gaze and not just the frequency of individual types of saccades, so that we could assess the extent to which the groups behaved differently across conditions. LDA is a biased classification technique that, when presented with any number of observations and a set number of classification groups, calculates the optimum separation between a combination of variables and reclassifies the observations into classification groups accordingly. Here, if the saccade patterns for each condition were distinct from those of other conditions, then the discriminant function would classify those data accurately; less distinctive patterns may be classified inaccurately. The LDA results indicate the percentage of individuals in a dataset for which the algorithm's classification matches the original conditions.

We elected to perform the LDA on the HP adults and HP children and not the LP children because the MANOVA between HP and LP children showed no significant difference between groups, which suggests that these groups would display a similar level of discrimination between conditions in an LDA. For this analysis, the classifiers for both HP adults and HP children were the five standardized AOI saccade types. These data were then classified into four different classification groups by the LDA, independent from the actual conditions. The classification groups were then compared against the original conditions in a confusion matrix, which displays how accurately the LDA algorithm's new classifications matched the original conditions (Table 3). The classification accuracy of each condition was calculated by dividing the number of correct classifications for each condition by the number of participants who had valid data for that condition.

The conditions of the twelve HP children were the most easily classified by the discriminant function, with an average classification accuracy of 56%. This was due to the high classification accuracies of the SD (83%) and SN (67%) conditions for the children. Notably, the saccade pattern of the SD and SN conditions were mutually exclusive for HP children, such that there were no cases in which the function confused one for the other, although it occasionally confused one for CO or IC. The function classified the CO data as SD or SN as often as it classified them correctly, for a CO accuracy of (33%). Similarly, the function classified IC data as SD or CO as often as it was correct, so the IC accuracy was 38%. Overall, each individual condition score was still greater than chance (25%), and the average classification accuracy (56%) was also above chance.

Adults had lower average classification accuracy than children, at 41%, The adults did not show clear differentiation for any of the four conditions, which was most notable in the relatively poor classification of SD (39%) and SN (48%) conditions. The CO condition was classified correctly 45% of the time and the IC condition, 31% of the time. The adults' saccade patterns for IC were more often confused with SD and the patterns for SD were more often classified as CO, which suggests that the adults focused more on the numerators of these problems, even though they ultimately responded correctly.

Table 3. Confusion matrices of original assigned conditions (y-axis) and the LDA predicted conditions (x-axis) for HP Adults and HP Children using the five standardized AOI saccade types as dependent variables. The correct classification accuracy for each condition is presented in the final column of each matrix as a percent. The overall correct classification accuracy is displayed in the bottom right corner of each matrix.

	HP Adults					HP Children				
	SD	SN	CO	IC	%	SD	SN	CO	IC	%
SD	13	4	11	5	39	10	0	1	1	83
SN	4	16	8	5	48	0	8	2	2	67
CO	7	6	15	5	45	3	4	4	1	33
IC	10	5	7	10	31	3	0	2	3	38
					41					56

Discussion

In this study we compared the mathematical strategy use of highly proficient children to that of highly proficient adults and less proficient children in order to investigate how age and proficiency independently contribute to mathematical achievement. We used a 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 cognitive strategies used during the task.

Our first hypothesis was that adults would demonstrate greater cognitive efficiency on the fraction comparison task than children, as measured by overall response times and total number of saccades/trial. To test this prediction, we compared HP adults and HP children, and found that HP adults did indeed exhibit faster RTs and fewer saccades than HP children. The groups had equally high accuracy scores on the task, which leads us to conclude that these behavioral differences are best explained by the development of cognitive efficiency as a function of maturity, as has been documented elsewhere (e.g., Luna et al., 2004). Further support for this claim comes from the comparison of HP to LP children, which showed no significant difference in measures of RT or number of saccades/trial, indicating that proficiency did not affect the efficiency with which children solved the problems.

Our second hypothesis was that HP children would demonstrate saccade patterns similar to those of HP adults and dissimilar to those of LP children and, more specifically, that participants who demonstrated high mathematical proficiency would correspondingly exhibit more task-relevant saccades than less proficient participants. Our rationale was that highly proficient participants would produce similar eye movements that reflected the use of similar strategies. This hypothesis was partially supported in our data, but also partially contradicted. Analysis of total saccades/trial supported the general hypothesis; both HP adults and HP children made more saccades/trial on difficult conditions than on simple conditions, showing responsiveness to task difficulty that the LP children did not show. This finding was also represented in the initial multivariate analysis of saccade types, wherein both HP adults and HP children showed an effect of task condition. In contrast to our hypothesis, these multivariate tests showed a significant difference between HP adults and HP children, and only a trending difference between HP children and LP children. We will first address the differences in the gaze patterns of HP adults and HP children, and then the differences between HP children and LP children.

HP adults and HP children showed a significant difference in their gaze patterns, primarily in NDl and NDx saccades. HP adults demonstrated more NDx saccades than HP children regardless of condition, and HP children demonstrated more NDl saccades than HP adults. Our interpretation of the higher prevalence of NDx saccades in the adult data is that adults were more likely to use a cross-multiplication strategy than were children who were just being introduced to fractions. The cross-multiplication algorithm requires the multiplication of each numerator by the opposing fraction's denominator, and the subsequent comparison of the products; it is an advanced procedure that may be taught in later grade levels. Although it is a mathematically valid shortcut to convert two fractions into fractions with equivalent denominators, the algorithm is often used in such a way that it is "devoid of meaning" (English & Halford, 1995). The prevalence of NDl transitions for children may indicate that they considered a greater number of numerical relationships than was necessary for the more experienced adult group. It may also be a byproduct of the HP children's overall greater number of saccades/trial, as noted previously.

Regarding task-relevant saccades, there were further differences between HP adults and HP children, contrary to our hypothesis that they would be similar. Specifically, we expected both HP adults and HP children to exhibit a greater number of between-numerator saccades on SD trials relative to other conditions, and a greater number of between-denominator saccades on SN trials relative to other conditions. This hypothesis was supported for HP children but not for HP adults. Whereas children showed the expected task-relevant saccade pattern for both the SD and SN conditions (Figs. 6 and 7), adults only displayed task-relevant saccades on the SN condition. Also contrary to our expectations, the LP children demonstrated task-relevant saccades on both the SD and SN conditions, just as the HP children did. This could mean that LP children recognize the patterns in the problems, but do not yet have a complete understanding of strategies or magnitudes. This result, in tandem with the difference in gaze patterns between HP adults and HP children, suggests that mathematical strategies are more influenced by maturity than by proficiency. However, these results could also be explained by a difference in general mathematical experience, in addition to cognitive maturity, and thus a comparison between novice adults and experienced children would be necessary to more accurately test this claim.

This particularly strong effect of age, instead of proficiency, for saccade patterns and task-relevant saccade use was unexpected and we offer several interpretations for these findings. First, although the HP children performed relatively well, with the exception of the IC condition, they were novices with fractions, and perhaps they needed to think through the problems methodically in order to arrive at the correct response, leading to exaggeratedly distinctive eye movements. The adults, who did not need to remind themselves of the rules, simply selected the correct answer with maximum efficiency. Second, it may be the case that the fraction problems, which were selected to be accessible to children, were simply too easy for adults. The minimal saccade differentiation per condition may indicate that adults used a simple scan strategy for every problem, and that their generally superior cognitive skills, including working memory, processing speed, and relational reasoning, allowed them to make a decision without making the additional saccades that the children did. A third possibility is that the screen layout allowed adults to use peripheral vision to a greater extent than did children. However, there is no indication in the literature that adults have superior peripheral vision to children.

There was no main effect of proficiency when comparing the HP and LP children. Figures 4 and 5 suggest that LP children responded more quickly and with fewer total saccades, specifically on the CO and IC conditions, but that test was not statistically supported. If that result were replicated with greater statistical power, it could indicate that LP children expended less effort on the more difficult problems by guessing quickly when confronted with relationships they did not understand. The data from this experiment do indicate that HP children not only understood how to solve most problems, but also showed persistence in solving the problems, particularly on the most difficult condition where their accuracy was the worst. This was evidenced by the main effect of condition on RT and saccades/trial in the analysis of HP children and adults, whereas these effects were not present in the analysis of HP and LP children. This mixed result indicates that the HP children showed a weak effect of responsiveness to condition difficulty while the LP children did not. Thus, more proficient children generally displayed a willingness to expend effort on solving the problem, regardless of understanding the numerical relationships.

Our final hypothesis was that people who had a greater understanding of fractions would show greater flexibility in their approach to problem-solving depending on the condition. We proposed that this would be manifested in unique patterns of saccades associated with each condition. However, HP adults showed a lower classification accuracy than the HP children, most likely because they made fewer eye movements overall. This finding does not preclude the possibility that adults may have adapted their strategies in a way that was not detectable with this paradigm.

Our findings contribute to the general understanding of the development of mathematical knowledge, and suggest related implications for pedagogy. We conclude that the cognitive abilities required for the fraction comparison task mature between middle childhood and adulthood, but are also affected by educational experience and task persistence. Children who have only begun to study fractions require more time and expend more cognitive effort than adults. A benefit of using eye-tracking to characterize development is the ability to identify the prevalence of specific strategies used at different time points. For example, in these data, only adults showed evidence of using the cross-multiplication strategy, which should not have been possible without former education on the cross multiplication strategy.

From these findings we suggest several implications regarding pedagogy. First, poor performance does not reflect the sophisticated task-relevant processing the LP children engaged in. Their saccade patterns for the SD and SN conditions indicate that they recognized the key features of each problem, but their incorrect responses suggest confusion with how to interpret those features. Because the HP children demonstrated similar gaze patterns, these findings suggest that most new learners approach these problems slowly and methodically, but that LP children lack the knowledge or confidence to draw the correct conclusion from mathematic stimuli at their stage of learning. One helpful pedagogical support could be to make all of the comparisons explicit and then state the relevant rule, to help make salient the numerical relations. For example, instead of simply reminding children of the rule (e.g., a smaller denominator indicates smaller pieces and therefore the same number of smaller pieces is a lesser fraction) it may be helpful to first point out the very basic featural comparisons (i.e., this denominator is smaller and the numerators are equal, therefore the denominator rule applies). We suggest this because the gaze data indicate that both groups of 5^{th}-graders focused on the key numerical features, and a recent review of pedagogical cognitive supports emphasized the importance of making basic relationships and comparisons explicit in order to facilitate higher level thinking (Richland et al., 2016).

Second, LP children did not demonstrate differences between conditions in total saccades or in RTs, whereas both the HP children and adults showed increased RTs and saccades/trial as task difficulty increased. Thus, although they directed their gaze toward the relevant information in each problem, the LP children either did not recognize how to handle the more difficult problems, or did not persist as long as the HP children did. In particular, the HP children persisted significantly longer on IC trials than the other conditions, presumably trying different strategies in an attempt to reconcile the conflicting information. In contrast, the LP children show a trend towards answering IC trials more quickly than the other conditions, which as stated before, could be the result of uncertainty and guessing. The current trend to incorporate growth mindsets (Dweck, 2006) in classrooms will be helpful as these children encounter more and more difficult concepts. Growth mindsets emphasize mastery, and the importance of learners' engagement in their own learning, with many programs encouraging struggle and even failure as opportunities for future learning.

Third, although the HP children performed nearly as well as the HP adults, they did not grasp the IC condition, which required them to integrate two conflicting pieces of information. It is our interpretation that this difficulty arises partly because of the relatively late maturation of the cognitive skill of relational integration. Developmental psychology abounds with instances in which very young children can proficiently integrate relationships and form analogies (Alexander et al., 1997; Gentner & Christie, 2014; Goswami, 1992), but an important contributor to success in those instances is a solid conceptual understanding of the domain (Goswami, 1992). In the context of fractions, this means an understanding of what fractions represent. These children were just beginning to work with fractions, so the comparison task presented here was at the edge of their conceptual reach. As they become more familiar with the fraction notation, the magnitudes associated with fractions, and more sophisticated strategies for handling them (e.g., converting to equivalent fractions, calculating decimals), they will be able to use this knowledge to resolve the conflicting relationships in the IC condition. However, it is our opinion that pedagogy could also help to highlight the relational nature of fractions, and make salient ways to integrate those relationships. Indeed, many adults, even after achieving computational proficiency with fractions, and after developing a mature ability to integrate relations in other domains, lack a relational concept of fractions (Stigler et al., 2010) and thus have an incomplete and ineffective understanding of mathematics in general.

The task conditions in this study spanned a fairly large range of difficulty levels for our novice learners. Future work in this arena could more carefully address an intermediate stage of fractions knowledge acquisition to look for pedagogical insights from eye-tracking. For example, as children make that conceptual leap from considering a single relationship between either numerators or denominators to simultaneously evaluating all four numbers, where does their attention fall and to what extent does that gaze behavior predict performance? Additionally, the reduction of data to correct trials primarily affected the statistical power in the analyses that included LP children, and may have impacted findings related to this group. Further research should seek to include more participants, or reconsider the interpretation of incorrect trials.

In the same way that this task was too difficult to glean further pedagogical insights for children, it was too easy to fully characterize adult proficiency. One caveat of this study is that, although this task was identical for all participants, the testing battery differed for adults and children; one of the other tests the adults completed was more difficult and is described in Miller Singley and Bunge (in prep). Furthermore, given the differing means of recruitment for children and adults – from a low-SES charter school versus a highly selective university – there are likely to be demographic differences between the two groups in addition to age and expertise.

In all, this study and similar prior research have established that eye-tracking can provide useful insights both for better characterizing the numerical cognition of proficient adults, and for illuminating the developmental and conceptual milestones in learners.

Chapter 2 – Strategy Adaptation in a Fraction Comparison Task: An eye-tracking study

Introduction

Everyday issues, both important and trivial, require that we reason about mathematical relationships. *If the United States government, employers, and citizens together spend 17.5% of the country's GDP on health care, and the Canadian government spends 11%, which system is the better deal, and for whom? 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? Should I support a local initiative that increases downtown parking rates to pay for the expansion of the light rail system, knowing it will impact my weekly budget in the short term but that I may be able to save money on my commute in the future?*

The link between mathematics and reasoning is apparent even 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 (C. T. Green, Mckerracher, Whitaker, Ferrer, & Bunge, 2011; Primi, Ferrão, & Almeida, 2010), and current educational policy promotes the use of analogical skills in a mathematics context (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 representation of rational numbers, they are inherently a mathematical concept, and yet, because they are defined by a numerical relationship, they have a relational reasoning aspect as well. Additionally, the conceptual difficulty associated with fractions leads to a wide variability in proficiency, persisting into adulthood, that has real consequences on learners' general academic and career success. It is not surprising that the topic of fraction understanding has garnered much attention both from cognitive psychologists and educational researchers in recent decades.

There are many conceptual challenges associated with learning fractions (Ni & Zhou, 2005; Siegler, 2016; Vamvakoussi & Vosniadou, 2004), and many students demonstrate lingering misconceptions. An infamous example comes from the 2004 National Assessment of Educational Progress, in which 50% of eighth graders failed to place the following fraction in increasing order by magnitude: 2/7, 1/12, and 5/9 (National Council of Teachers of Mathematics, 2007). Misconceptions at lower grades turn into obstacles in advanced mathematics topics such as algebra, and can even persist into adulthood (DeWolf et al., 2014; Stigler et al., 2010). It has been established that fractions knowledge is a strong predictor of general mathematics performance (Booth, Newton, & Twiss-Garrity, 2014; Geary, Hoard, Nugent, & Bailey, 2013; Schneider, Grabner, Zurich, & Paetsch, 2009; Siegler, Thompson, & Schneider, 2011), particularly of readiness for algebra (Booth & Newton, 2012). Thus, fractions knowledge is a stepping stone to success in STEM careers – and, because Algebra I is a high school graduation requirement in 25 American states (Reys, Dingman, Nevels, & Teuscher, 2007), failure to learn fractions can constrain learners' general academic success and limit their potential career options. Clearly not everyone is subject to these misconceptions, however - after all, 50% of eighth graders answered the above question correctly. The National Center for Education Statistics reported that 40% of bachelors degrees conferred to men in 2014 were in STEM fields, and 29% for graduating women (U.S. Department of Education. Institute of Education Sciences. National Center for Education Statistics, 2015). Thus, there is wide variability of fractions knowledge exhibited in the American public.

One of the key indicators of fractions competence is the ability to translate the formal *a/b* notation into a holistic magnitude, as tested in the NAEP exam question. The failsafe method to assess the magnitude associated with a fraction is to calculate the decimal value of the numerator divided by the denominator, but this is cumbersome and prone to error. If the fraction's components make it feasible, one could alternatively calculate the relationship of the numerator to the denominator, or convert to an equivalent fraction that makes the relationship more salient. For example, in the fraction 12/16, most schoolchildren are taught to divide both numerator and denominator by the common factor 4 to convert to the equivalent fraction 3/4, for which the magnitude is easily recognizable as 0.75. For the fraction 13/16, that conversion is not possible, but a numerically-confident learner may be able to estimate that magnitude as being a bit larger than 0.75.

Although there are various strategies for assessing fraction magnitudes, it is clear that the ability to do so is critical to mathematical proficiency. Siegler, Thompson, and Schneider (2011) found that 11-13 year-olds' skill at placing fractions on a number line according to their holistic magnitude was correlated both with their computational skill with fractions as well as their overall mathematics achievement scores. Booth and Newton (2012), and Booth, Newton, and Twiss-Garrity (2014) related fraction magnitude placement directly to measures of algebraic readiness and algebraic equation-solving in 12-15 year olds.

Fraction Comparison Task

A different way to test fraction magnitude knowledge, aside from the number line placement paradigm, is the fraction comparison task, in which two fractions are displayed side by side and participants are asked to select the larger (or smaller) one (Clarke & Roche, 2009; Faulkenberry & Pierce, 2011; Fazio et al., 2016; Gabriel et al., 2013; Ischebeck et al., 2009, 2016, Meert et al., 2009, 2010a, 2010b; Obersteiner & Tumpek, 2016; Smith III, 1995). This task has been shown to elicit a variety of strategies from purely mathematical algorithms to complex relational comparisons (Clarke & Roche, 2009; Fazio et al., 2016; Smith III, 1995), which makes it an ideal setting for investigating the links between relational reasoning and mathematical cognition.

Prior research using the fraction comparison task has made inferences about the nature of mental representations, specifically whether people tend to mentally represent fractions by their holistic magnitudes, referred to as the holistic processing theory, or by their individual components and the relationships between them, called the componential processing theory. The distance effect 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), therefore, manipulating the distance between the components versus the distance between the fraction magnitudes gives 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 et al., 2009) and holistic magnitude processing (Faulkenberry & Pierce, 2011; Ischebeck et al., 2009; Sprute & Temple, 2011). Quite a few additional studies have shown behavior indicative of both componential and holistic processing, depending on the task context (Fazio et al., 2016; Gabriel et al., 2013; Meert et al., 2009, 2010a; Smith III, 1995), suggesting a hybrid mental representation.

Probing fraction comparison strategies via eyetracking

Chronometric studies, as described above, enable inferences about the mental representations underlying fractions; however, cognitive processing strategies 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 et al., 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 strategies as they evaluate fraction magnitudes.

The few eye-tracking studies to date have supported the hybrid theories of fraction representation promoted by the initial behavioral 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 & Tumpek (2016) reported similar results, and additionally found that people exhibited more behavior consistent with magnitude processing (i.e., more numerator-denominator saccades) when looking at mixed pairs than when they were looking at pairs with common components. Taken together, these studies provide further evidence for the hybrid model, or the theory that people adaptively switch between mental representations depending on task demands.

However, both the behavioral and eye-tracking results described above depend on the fraction pairs selected for comparison, and as such the 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 or 5/9 vs 8/9, it encourages componential processing because only the non-identical numbers need to be compared. Conversely, using over-learned or familiar fractions, such as 1/4 vs 3/4 or 2/5 vs 7/10, makes it easier to assess the holistic magnitude of the fraction. The finding that people are sensitive to these manipulations and adjust their strategy accordingly constitutes strong evidence for the hybrid representation model – and some recent findings have linked strategy flexibility to overall mathematical proficiency (Fazio et al., 2016; Smith III, 1995). However, it is an open question whether both componential and holistic representations, or even flexibly switching between them, are beneficial absent the availability of heuristics.

Moreover, the studies including mixed pairs as well as pairs with common components generally concluded that mixed pairs were more difficult, required more saccades overall, and people exhibited more vertical saccades than for the pairs with common components. These findings imply that people consistently use holistic magnitude approaches for mixed pairs. To examine strategy use and adaptation specifically for mixed pairs, we tracked saccades during more subtle manipulations to investigate whether and to what extent people use componential versus holistic processing.

Both componential and holistic processing may be adaptive in more subtle task manipulations. For example, the pair 2/3 versus 5/12 could be solved either by estimating that 5/12 is less than 0.5 while 2/3 is greater, or by noting that the right denominator (12) is 4 times the left (3) while the right numerator (5) is only 2.5 times the left (2). Both strategies – comparing estimated magnitudes and comparing componential relationships – leads to the answer that 2/3 is greater than 5/12. The specific numbers involved in a pair will determine whether these approaches are equally feasible.

Fraction Comparison task in a Relational Reasoning Framework

The terms componential and holistic processing characterize the fraction comparison task as a purely mathematical problem. Based on our earlier assertion that fractions are an ideal setting to investigate the interaction of mathematical and relational reasoning, we find it informative to additionally characterize the fraction comparison task in relational reasoning terms. 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. 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. In relational reasoning terms, this is a second-order comparison of two first-order relationships (Miller Singley & Bunge, 2014). Highlighting the relational reasoning aspect of this mathematical task allows us to investigate the relationship between relational reasoning in a mathematical context and relational reasoning in an abstract, visuospatial context.

Research Questions and Hypotheses

In the present study we investigated the relational and mathematical strategies associated with a difficult version of the fraction comparison task, and their impact on performance. We selected two manipulations that we expected to facilitate opposite processing approaches – between-fraction/componential or within-fraction/holistic comparisons – and tested the participants' adaptiveness and resulting 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. 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 used a ratio of between-fraction saccades to between-plus-within-fraction saccades to investigate the extent to which the different types of relational comparisons contributed to performance.

The distance effect, or the phenomenon that it is easier to judge the magnitudes of numbers that are farther apart, is known to affect performance on the fraction comparison task (Schneider & Siegler, 2010; Sprute & Temple, 2011). For example, it is easier to judge 2/11 vs 8/9 than it is to judge 3/5 vs 4/7. Although there is some evidence that this effect is influenced by the distance between components (Ischebeck et al., 2009), it is also influenced by the fraction magnitude itself (Ischebeck et al., 2009; Sprute & Temple, 2011), and as such is taken as evidence for the holistic processing theory. Because we wanted to probe the effects of more subtle manipulations, we restricted the range of magnitude difference in order to constrain its overall effect.

We hypothesized that participants would adapt their gaze behavior to demonstrate more horizontal saccades for trials with multiples and more vertical saccades on trials with magnitudes opposite 1/2, relative to the individuals' overall average behavior. Secondarily, we hypothesized that these gaze adaptations would be associated with better performance, such that the extent to which an individual adapted their gaze behavior in response to the trial condition would predict their performance on that trial. As an extension of this hypothesis, we also tested whether participants who more consistently used one strategy performed better overall. Because Obersteiner & Tumpek (2016) found a greater proportion of vertical than horizontal saccades in their sample of skilled mathematicians, we hypothesized that those participants who demonstrated a greater proportion of vertical saccades would have better overall performance than those who demonstrated a greater proportion of horizontal saccades. Finally, we tested for correlations between participants' mathematical performance and their performance on a visuospatial relational reasoning task, with the hypothesis that these constructs would be related.

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). Participant recruitment and 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 testing 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. The first block, which was self-paced and lasted 2-5 minutes, was a variation of this fraction comparison task, adapted from Ischebeck, Schocke, and Delazer (2009). It included fraction pairs with the same numerator or the same denominator, along with pairs that shared no common components. This first testing block was acquired as a point of comparison for a developmental study reported elsewhere (Crawford, Miller Singley, & Bunge, in prep). The second two blocks, each lasting approximately 3-5 minutes, formed the basis of the present investigation. To provide a break, the blocks were separated by the Analysis-Synthesis sub-test of the Woodcock Johnson III Tests of Cognitive Abilities (Woodcock, McGrew, & Mather, 2001). After that paper test, which is an independent measure of domain-general relational reasoning, they returned to the eye-tracker to complete the second block of this fraction comparison task. Two final blocks were additional variants of the fraction comparison task: one using improper and proper fractions in which each pair had either the same numerators or same denominators, and one block containing one sample 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); two participants did not meet this criterion on the two blocks included in this study and were excluded from analyses. Five additional participants' performance was not different from chance (50% correct overall; based on one-sample t-tests, p ranged from 0.11 to 0.37), so they were also excluded. For the gaze analyses, we evaluated only the trials that included specific saccades of interest: those between the four digits on the screen. One participant's average accuracy on the trials that contained saccades of interest 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 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 2x2 factorial experimental design, described below (Fig. 1). The two factors were selected because they promote specific and opposite problem-solving strategies, allowing us to investigate whether and how people aligned their problem-solving strategies to the affordances of the stimuli, and how that alignment affected their performance. The four conditions described below were interspersed pseudorandomly throughout the two task blocks. The first three participants had 10 seconds to complete each trial, but we found they completed the trials much more quickly and so we adjusted the time limit to 4 seconds for the remaining participants. For the first three participants, we recoded as incorrect any of their correct responses that were made after 4 seconds.

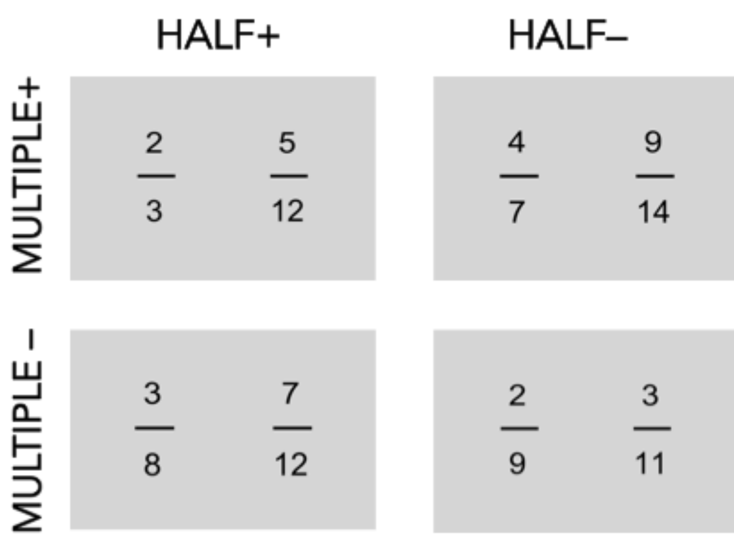


Fig. 1. Experimental design and sample problems. Half+ problems included one fraction with a magnitude larger than 0.5 and one with a magnitude smaller. Half– problems included two fractions that were both either larger or smaller than 0.5. Multiple+ problems included one denominator that was a multiple of the other whereas Multiple– problems did not.

Factor One (Multiple) 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., 4/5 vs 7/10), referred to subsequently as M+, whereas the other half did not (e.g., 4/5 vs 7/8; referred to as M-). 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, 4/5 * 2/2 = 8/10, which is an easy comparison to 7/10. 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 individual components of the two fractions. Attempting to use the same strategy on M- trials, such as 4/5 vs 7/8, 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.

Factor Two (Half) 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 promoted the use of the benchmarking to 1/2 strategy, in which one can quickly estimate or calculate whether each fraction's magnitude is greater or less than one half, and then make a simple numerical comparison between the two magnitudes. We consider this strategy to reflect holistic magnitude processing because the focus of attention is on the relationship between the numerator and denominator within each fraction. For H- trials, wherein the fraction magnitudes are both either greater than or less than 1/2 (e.g., 3/7 vs 4/9), 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-. The first condition facilitated the use of either equivalent fractions or benchmarking strategies, and we predicted that participants would perform the best on this condition. The M+H- condition was designed to promote the equivalent fractions strategy, and therefore we expected that participants would first convert denominators and then compare numerators on these trials. The most efficient saccade pattern would consist of one horizontal saccade to compare the denominators, and another to compare the numerators. There is necessarily one vertical or diagonal saccade to move between top and bottom of the screen, but an emphasis on between-fraction comparisons would result in a greater proportion of horizontal saccades relative to other types of saccades. The M-H+ condition was designed to promote benchmarking, and therefore we expected that participants would estimate the magnitude of each fraction, then compare the resulting magnitudes. This strategy would manifest as one vertical saccade to make each within-fraction comparison or magnitude estimate, and one horizontal to move between the fractions, leading to a greater proportion of vertical eye movements. Finally, the M-H- condition obstructed both strategies, so we expected worse performance on these trials than on the others. If participants are familiar with the cross-multiplication strategy, they might resort to that on M-H- trials because it can be applied to any pair. However, it is slower than the other heuristics and more difficult. If they do cross-multiply, we would expect to see two diagonal saccades between the fractions, and either a vertical or horizontal saccade as their eyes switch between components. In fact, however, diagonal saccades made up only 5%

of all recorded saccades, and there was no indication that there were more diagonal saccades in the M-H- condition than the other conditions.

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, with larger numerators suggesting larger fractions and larger denominators suggesting smaller fractions. Thus, participants had to take all four digits 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 0.01-0.27. Another potential distractor is the size of the numerical components. Although 2/3 represents a greater magnitude than 12/19, it is tempting to assume that larger numbers comprise the larger fraction. Therefore, we ensured that half of the larger fractions were composed of smaller numbers. Finally, all fractions were proper (between 0 and 1), irreducible, composed of digits 2-19, 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 the Appendix.

Data collection

The experiment was conducted on a Tobii T120 eye-tracker, with a sampling rate of 120Hz (or every 8.3ms). Participants sat approximately 64cm from the eye-tracker, per Tobii specifications. Each block of the task 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 one second between trials. During trials, numbers were displayed in black sans-serif text on a grey background and were 2.7cm in height and 1.9cm in width if the number was comprised of one digit, or 3.8cm for two digits. There was a 2.7cm vertical distance between the numbers and the fraction bar, and a horizontal distance of 14.6cm 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 coveting 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 two fraction bars and the three corresponding areas between the two fractions (Figure 2). 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 if need be. 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 100-500 milliseconds (Holmqvist et al., 2011). Any fixation that lasted less than 40 milliseconds was presumed to be a transit between AOIs instead of a true fixation, and thus was excluded from further analysis. Any contiguous fixations in the same AOI that were separated by fewer than 100 milliseconds of missing samples were concatenated, as the disruption was presumed to be caused by a blink.

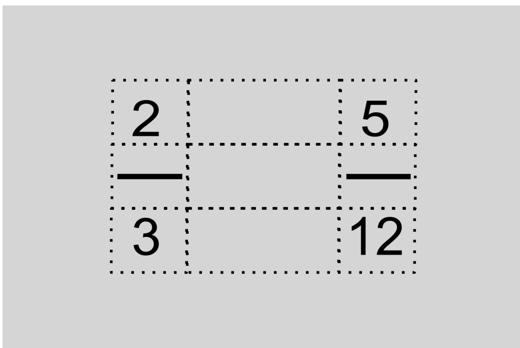


Fig. 2. Areas of Interest (AOIs) were defined within the experiment. The eye-tracker designated which AOI, if any, was the focus of gaze for each sample.

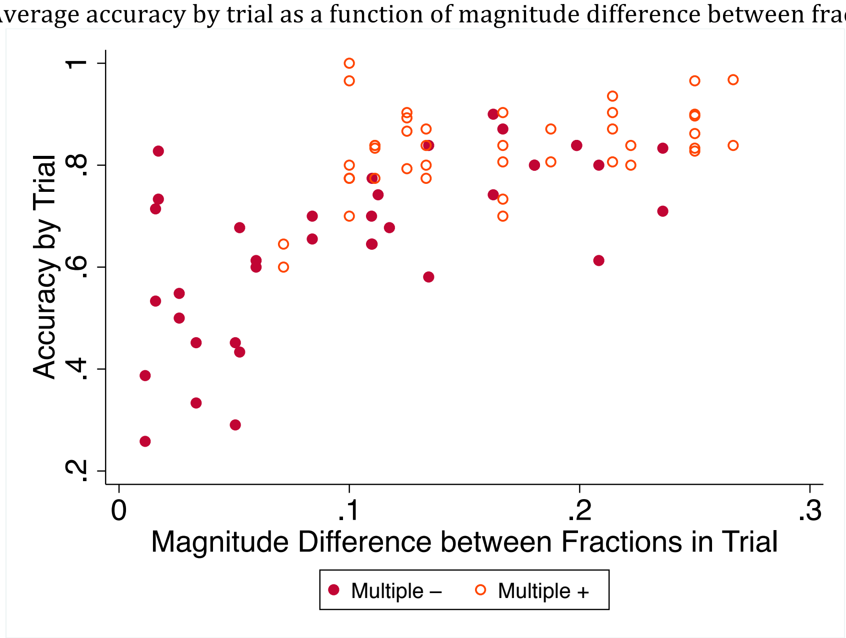
Because our research questions pertained to differential patterns of eye movements, our primary metric was the number of saccades from one AOI directly to another within each trial. Using the list of fixations generated as described above, we counted the number of saccades between each of the four primary AOIs. 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. Any two fixations that were separated by more than 300ms, the duration of a typical blink (Holmqvist et al., 2011), were not counted as a saccade, because it is possible the eyes moved to a third location in that timeframe instead of making a direct saccade. To collapse the horizontal and vertical metrics into a single metric characterizing each trial, we calculated the ratio of horizontal to total saccades as follows: horizontal / (horizontal + vertical). This metric approaches 1 when there are more horizontal than vertical saccades in a trial, and 0 when there are more vertical than horizontal. As stated above, the most efficient saccade ratio when converting to equivalent fractions is 0.67. Additional horizontal saccades would increase the ratio and, in our interpretation, reflect more focused, albeit less efficient, between-fraction comparisons. The most efficient saccade ratio when benchmarking to 1/2 is 0.33, with additional vertical saccades decreasing the ratio. Our inference is that a lower ratio suggests participants are paying greater attention to within-fraction comparisons. Ratios in the middle indicate either use of different strategies or general confusion. All additional 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 number of saccades per trial was 6.43 (SD=3.06, range 0-30), and the subset of horizontal and vertical saccades per trial averaged 2.56 (SD = 2.43, range 0-15). All trials were included in the analysis of the conditions' effect on behavior, but trials with no recorded horizontal or vertical saccades (22.7% of the total 2,344) were excluded for saccade-related analyses, leaving 1,811 trials with an average combined count of horizontal and vertical saccades of 3.29 (SD = 2.27). Trials on which participants did not respond within the 4-second response window were marked as incorrect.

Analytic approach

Based on prior research indicating that magnitude difference impacts difficulty of fraction comparisons (Sprute & Temple), 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 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 (Fig. 3). 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 many of the M- trials also had smaller magnitude differences (Fig. 3), 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 behavioral effects of the task conditions, we used regression models at the trial level that included both Multiple and Half factors as well as a continuous factor of Magnitude Difference.

Fig. 3. Average accuracy by trial as a function of magnitude difference between fractions in the pair.



Per our hypotheses, we tested whether behavioral performance was sensitive to the task conditions by conducting a mixed-level logistic regression on accuracy (0 or 1) per trial. The two categorical trial factors and the continuous magnitude difference between fractions were entered as predictors, as well as a random effect of participant to account for participant-level dependence across trials. To explore the conditions' effect on response times and on total number of saccades per trial, we conducted mixed-level regressions with the same predictors as above. In the following regression model, Y alternately represents accuracy, response time, and all saccades in three independent analyses. B_0 represents the predicted intercept, and B_1-B_4 represent the regression coefficients corresponding to the Multiple, Half, Magnitude Difference, and Multiple by Half interaction effects, respectively. Zeta represents the participant-level cluster effect, and Epsilon represents trial-level error.

$$Y_{ij} = B_0 + B_1{*}Mult_{ij} + B_2{*}Half_{ij} + B_3{*}MagDiff_{ij} + B_4{*}Mult{*}Half_{ij} + Zeta_j + Epsilon_{ij}$$

All subsequent analyses were conducted only on the trials that contained the primary saccades of interest: horizontal or vertical between digits, using the horizontal/(horizontal + vertical) saccade ratio as the primary metric. The use of a cross-multiplication strategy would be expected to result in diagonal saccades, which we predicted would be most common on the M-H- trials, when the other heuristic strategies were unavailable. However, only 5% of all recorded saccades were diagonal, and there was no indication that there were more diagonal saccades in the M-H- condition, nor that they conferred any performance benefits. Therefore, we proceeded with the saccade ratio metric as defined by horizontal and vertical saccades.

Results

Task performance

Our first analyses tested our predictions that M+ and H+ trials would yield better accuracy than M- and H- trials. The mixed logistic regression yielded a predicted accuracy of 81.5% for M+H+ trials, calculated at the mean value of Magnitude Difference, ($Z=9.43$, $p<0.001$; Table 1). M-H+ trials were associated with 10% lower accuracy, estimated at 71.3%, ($Z=-3.59$, $p<0.001$). There was no main effect of Half; predicted accuracy on M+H- trials was 83.2%, ($Z=0.73$, $p=0.50$). The interaction of the two factors was also not significant: predicted accuracy of M-H- trials was 66.6%, ($Z=-1.61$, $p=0.10$), 15% lower than M+H+ trials. Increasing the distance between the fractions' magnitudes by 0.01 was associated with a 6.2% increase in the predicted accuracy, ($Z=5.66$, $p<0.001$). 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 2153ms for M+H+ trials, calculated at the mean value of Magnitude Difference, ($Z=28.39$, $p<0.001$). Both Multiple and Half factors exhibited main effects: RTs for M-H+ trials were estimated to be 195ms slower than for M+H+ trials, ($Z=4.44$, $p<0.001$), and RTs for M+H- trials were 186ms faster, ($Z=-4.35$, $p<0.001$). The interaction was also significant such that M-H- trials were 243ms slower than M+H+ trials, ($Z=3.87$, $p<0.001$). Additionally, a 0.01 increase in magnitude difference was associated with a 16ms faster response, ($Z=-5.46$, $p<0.001$). 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 were faster 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.

To investigate whether the task conditions impacted the overall number of saccades per trial, we used the same type of mixed regression model as for the behavioral measures. This model yielded a predicted total number of saccades of 6.18 for M+H+ trials, calculated at the mean value of Magnitude Difference, ($Z = 24.59$, $p < 0.001$). There was a main effect of Multiple such that the M- trials had 0.96 more saccades, ($Z = 5.79$, $p < 0.001$). There was a main effect of Half such that the H- trials had 0.67 fewer saccades, ($Z = -4.03$, $p < 0.001$). There was also an interaction of Multiple and Half such that M-H- trials had 0.50 more saccades, ($Z = 2.20$, $p = 0.028$). Magnitude Difference also impacted total number of saccades: a 0.01 increase in magnitude difference was associated with a predicted 0.06 decrease in saccades per trial ($Z = -5.13$, $p < 0.001$). These results reflect those for RTs; all three manipulations affected the extent to which participants analyzed the stimuli prior to making a magnitude judgment.

Table 1. Results of regression analyses testing effects of conditions on behavior and gaze.

Factor	Accuracy (Z)	Response Time (Z)	Total Saccades (Z)	Saccade Ratio (Z)
Multiple –	-3.59***	4.44***	5.79***	-2.28*
Half –	0.67	-4.35***	-4.03***	0.58
Multiple – x Half –	-1.61	3.87***	2.20*	1.13
Magnitude Difference	5.66***	-5.46***	-5.13***	0.76
Constant (M+H+)	9.43***	28.39***	24.59***	13.82***
***Participant-level Effect* (ρ)**				
	0.07**	0.26***	0.16**	0.31*

Asterisks indicate p-values: * = p<.05; ** = p<.01; *** = p<.001

Gaze Analyses

We next turned to our investigation of how participants adapted their strategies based on trial type by testing for an effect of condition on the saccade ratio (the proportion of horizontal saccades relative to both horizontal and vertical saccades). The overall mean saccade ratio across all trials is 0.54, with a standard deviation of 0.36, reflecting a tri-modal distribution rather than a large number of trials with a 1:1 ratio. The mixed regression model showed that horizontal saccades were slightly more prevalent than vertical across participants, with the overall ratio for M+H+ trials at 0.56, ($Z=13.82$, $p<0.001$; Fig. 4). Multiple had a main effect on the saccade ratio, with 4.9% more horizontal saccades relative to vertical saccades on M+H+ than M-H+ trials ($Z=2.28$, $p=0.023$). This result aligns with our hypothesis that people would use more between-fraction comparisons – and specifically the strategy of converting to equivalent fractions – on M+ trials, in which one denominator is a multiple of the other. Note that the random effect of participant in the model accounts for individuals' average range of saccade ratios; therefore, this main effect of Multiple reflects an overall adaptation to the stimuli exhibited by many individuals in the sample. There was no effect of Half, Multiple by Half interaction, or Magnitude Difference on the saccade ratio. This analysis was conducted on all trials that contained horizontal and/or vertical saccades, and the results were slightly stronger when restricting the analysis to correct trials only. The prevalence of horizontal saccades suggests that participants were more likely to attempt the equivalent fractions strategy, a result that aligns with the behavioral results that participants were faster and more accurate on M+ trials. It does not appear that participants made any more within-fraction comparisons on H+ than H- trials, even though we expected H+ trials would favor magnitude processing.

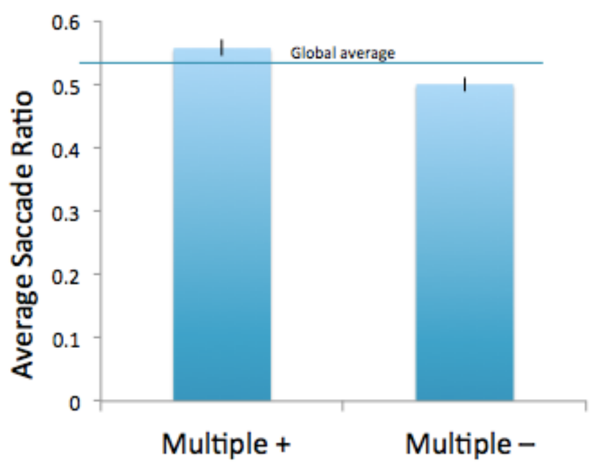


Fig. 4. Participants made a greater proportion of horizontal saccades on M+ than on M– trials.

Our final set of analyses investigated whether adaptations in gaze strategy impacted performance, both within and between participants. Our predictions were that a greater proportion of horizontal saccades would lead to better performance – both faster and more accurate – on M+ trials, and that a greater proportion of vertical saccades would lead to better performance on H+ trials. We conducted these analyses within-participant, to test for advantages of flexibility, as well as between-participant, to test for overall advantages of one strategy versus another.

First, we tested for within-participant performance differences based on the saccade ratio. A mixed-level regression at the trial level testing the effect of saccade ratio, including a fixed effect of participant, showed no difference in accuracy due to saccade ratio differences between trials, (Z = 1.08, p = 0.28). The same regression model conducted on RTs showed that a greater proportion of horizontal saccades was associated with faster RTs overall, (Z = -2.84, p = 0.005) – and this was the case even when controlling for task condition, (Z = -2.16, p = 0.03). Follow-up regression analyses exploring the differential effects on RTs for M+ and M- trials confirmed that the benefit of horizontal saccades was limited to M+ trials, wherein a greater proportion of horizontal saccades was associated with faster RTs (t = -2.06, p = 0.04) when controlling for Half (t = -2.75, p = 0.006) and Magnitude Difference (t = -3.19, p = 0.002). None of these factors have significant effects on RTs for M- trials. Therefore, a greater proportion of horizontal saccades conferred performance benefits relative to an individual's average RT, even after accounting for difficulty of task condition – particularly for M+ trials. There were no performance benefits associated with a greater proportion of vertical saccades, despite our prediction that vertical saccades would facilitate magnitude processing and thereby boost performance on H+ trials.

Next we tested for differences between participants based on their average saccade ratio over all trials (Table 2). A linear regression of average accuracy on average saccade ratio did not explain any differences between participants on accuracy, (t = 0.97, p = 0.34). Like the within-person analyses above, a greater average proportion of horizontal saccades was associated with faster average RTs on correct trials, (t = -2.25, p = 0.03). However, the addition of condition in a mixed regression analysis rendered non-significant the effect of saccade ratio on average RTs (Z = -0.58, p = 0.56), because the faster RT averages were associated with M+ trials (t(114) = -4.85, p < 0.001). Although the people who had higher overall average proportions of horizontal saccades also had faster overall RTs than their peers, the differences in condition outweighed the effect of saccade ratio.

Table 2. Results of regression analyses testing effects of saccade ratio on performance.

Factor	**Average Accuracy (t)**	**Average RT (Z)**	**Average RT (Z)**
Saccade Ratio	0.97	-2.25*	-0.14
Multiple –	–	–	4.02***
Half –	–	–	-1.20
Magnitude Difference	–	–	1.89
Constant (M+H+)	12.94***	13.83***	15.71***
***Participant-level Effect* (ρ)**			
	–	–	0.61***

Asterisks indicate p-values: * = p<.05; ** = p<.01; *** = p<.001

Our final question was whether participants' saccade ratio was associated with their score on our measure of non-numerical reasoning, the Analysis-Synthesis paper test. A Pearson's correlation of Analysis-Synthesis score with average saccade ratio for each condition, as well as average over all trials, yielded null results, (overall: r=0.1, p=0.59; by conditions: r ranges from -0.13 to 0.27, p ranges from 0.15 to 0.94). While it is possible that relational thinking provides a foundation for mathematical thinking, this version of the fraction comparison task, which uses larger numbers and smaller magnitude differences than other versions, seems to be more sensitive to individual differences in computational skill than in relational reasoning.

Discussion

We investigated adults' strategies on a fraction comparison task, and specifically whether attention to within-fraction or between-fraction relationships enhanced mathematical performance. The fraction pairs were designed to promote 1) between-fraction comparisons, also known in the literature as componential processing, via the use of an equivalent fractions strategy, or 2) within-fractions comparisons, known as holistic magnitude processing, via a benchmarking to 1/2 strategy. We measured strategy use by calculating the relative number of horizontal eye movements, an indicator of componential processing, to vertical eye movements, an indicator of holistic processing, on each trial. We then tested whether participants adjusted their strategies according to task condition, and compared strategy use both within and between participants to investigate whether one approach or the other was more adaptive.

First, we found that participants exhibited better accuracy, faster RTs, and fewer total saccades on trials in which one denominator was a multiple of the other (M+). In this task condition, it was easy to identify the factor that could be used to convert one fraction to an equivalent fraction with the same denominator as the other, which facilitates direct comparison of the numerators. This strategy is called the equivalent fractions strategy, and is associated with horizontal saccades, as people direct their gaze first between denominators and then between numerators. Accordingly, we also found an increased proportion of horizontal saccades on these same trials. Taken together with the finding of fewer overall saccades on M+ trials, we have evidence that people took advantage of the optimal strategy for M+ trials and were more efficient when they did so.

This finding extends the findings of previous studies, in which strategy adaptation was reported, although these prior studies primarily addressed broad differences between pairs that share common components (e.g., have identical numerators: 2/5 vs 2/7) versus those that do not. Both Ischebeck, Weilharter, and Körner (2016), and Obersteiner and Tumpek (2016) found that participants' saccades indicated componential processing when fraction pairs shared common components, and holistic processing when fraction pairs did not share common components. Ischebeck and colleagues recorded more saccades between numerators when the pair shared an identical denominator, and vice versa, indicating that people used componential processing when the fractions shared common components. They also noted that pairs without common components led to a greater overall number of saccades, indicating that the increased processing demand was manifested in increased saccades. Additionally, Obersteiner and Tumpek reported an increase specifically in vertical saccades for pairs without common components, indicating preferential use of holistic processing. In our study, when there were no common components between any fraction pairs, we still found that people adjusted their strategies in response to the task condition, by attending to different types of relations, even though the conditions were relatively subtle.

Moreover, the findings from the current study contrast somewhat with these previous reports; we found evidence of componential processing even among fraction pairs that did not share common components. Specifically, our participants made a greater proportion of horizontal saccades on trials in which between-fractions comparisons were strategically beneficial.

In another extension of previous work, we investigated the effect of saccades on accuracy and RT, both within and between participants. Although there was no difference in accuracy as a result of saccade ratio, we found RTs to be faster on M+ trials with a greater proportion of horizontal saccades. This was true within participants even when controlling for task difficulty, and also when comparing participants' average RTs relative to their peers. In particular, and in line with our hypothesis, a greater proportion of horizontal saccades was associated with faster RTs on the trials that promoted the equivalent fractions strategy. Thus, not only did people appropriately adjust their strategies, but doing so improved their performance.

We had predicted that flexibility in strategies would not only lead to better performance on this fraction comparison task, but also reflect greater skill in non-numerical relational reasoning. However, participants' scores on the non-numerical reasoning task were unrelated to their fraction comparison performance. It remains possible that non-numerical and mathematical reasoning share a common relational mechanism, but that we did not detect it with these tasks. We suspect that, because this mathematical task relied heavily on participants' computational skills, it was less sensitive to differences in relational reasoning. This hypothesis could be investigated with better reliability in an experiment that manipulates relational reasoning within a mathematical context.

Our findings regarding improved performance were limited to the beneficial effect of horizontal saccades on M+ trials. We had also predicted that a greater proportion of vertical saccades would be associated with better performance on H+ trials, but our data did not support that prediction. Despite a strong effect of Magnitude Difference on behavior, which implies that people were attending to the fractions' magnitudes, we did not evidence that magnitude processing was helpful. Instead, we saw a greater proportion of vertical saccades on M- trials (that obstructed the equivalent fractions strategy), possibly indicating that people attempted to use the equivalent fractions strategy and, when it was difficult to do so, they attempted other strategies. Additionally, the association between faster RTs and a greater proportion of horizontal saccades supports the interpretation that participants first attempted an equivalent fractions strategy, which was adaptive only on certain trials. If that strategy didn't work, their attempt to use a different strategy cost them time. Therefore, it is possible that people simply did not notice the opportunity to benchmark to 1/2, and fixated on the strategy of converting to equivalent fractions, even when it was not adaptive.

Another possible explanation for the prevalence of horizontal saccades is that this task was second in the testing session. The participants completed an initial fraction comparison task that did include trials with common components. Therefore it is possible they were used to making between-fraction comparisons and simply continued to do so on this task. However, if that were the case, one would expect the second block of this task to contain fewer horizontal saccades, but there was no indication of block order effect on horizontal or vertical saccades on the task presented here.

A limitation of this analysis is that our conservative accounting of saccades contributed to a relatively low per-trial average. We 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 300ms. A second potential source of the relatively low saccades per trial is the fact that we limited participants' RTs to 4 seconds in order to motivate efficient processing. However, it is possible that placing time pressure on participants motivated them to stick with a particular strategy instead of taking the time to choose the optimal strategy for each trial.

Another limitation is that we did not collect survey data about which mathematical strategies the participants remember learning from their elementary curriculum. It is possible our participants had initially learned or currently use a wide variety of strategies to manipulate fractions, while we tested for relatively few.

As such, this study is a first indication that people adapt their fraction comparison strategies even within subtle task manipulations. It extends previous findings by showing that participants attended to different types of relations, within-fraction or between-fraction, depending on the trial affordances, and doing so enhanced their task performance. This demonstration of participants' strategy adaptation within trials adds mechanistic insight to prior findings that strategy shifting is associated with better overall mathematical performance. This may be an initial indication that relational reasoning, and particularly the ability to identify multiple dimensions of relationships within a single problem, forms a foundation for mathematical proficiency. As such, this research raises further inquiries regarding the development of fraction proficiency, as well as the source and extent of individual differences among more heterogeneous samples.

Section Two

Chapter 3 – Curriculum, Classroom, and Individual Effects on Students' Mathematics Learning Trajectories

AERA Conference proceedings prepared by Alison T. Miller Singley (first author), Nicole Leveille Buchanan (second author), and Chloe T. Green (third author); presented April 7, 2014 in Philadelphia, PA.

Acknowledgements

The research reported here was supported in part by the Institute of Education Sciences grants R305B070299 (research grant) to Geoffrey B. Saxe and R305B090026 (predoctoral training grant) to the University of California, Berkeley. The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences or the U.S. Department of Education. We are grateful to the teachers who participated in our research, and to the many people who contributed to various phases of the Learning Mathematics through Representations (LMR) project.

Introduction

It is increasingly common for commercial mathematics curricula to incorporate research on learning trajectories, or "the levels or waypoints of thinking, knowledge, and skill in using knowledge, that students are likely to go through as they learn mathematics" (Daro, Mosher, & Corcoran, 2011). Research on learning trajectories has focused on identifying optimal sequences of learning activities through which students are best supported in constructing generative understandings of various topics. Key advancements in this field have recently been incorporated in standardized policy (Common Core State Standards Initiative, 2010). However, some authors point out that social and individual factors may have great influence in shaping students' actual learning trajectories, affecting how students respond to sequenced lessons in curriculum materials (Battista, 2011; Empson, 2011). Currently, there is little empirical evidence of individual and classroom/social factors influencing students' learning trajectories relative to the effects of the sequence of topics embedded in curricular materials. In other words, as the question was posed by Battista (2011), "If one constructs a prototypical hypothetical LT [Learning Trajectory] for a particular topic, how do the actual LT for individual students vary about this prototypical path?" (p. 61). We posit that individual prior learning experience and classroom social factors are likely to significantly impact students' actual learning trajectories, and moreover, a lack of understanding of these factors may lead researchers and teachers to have unrealistic expectations for student progress along a given curriculum's learning trajectory.

Our research aims to unify existing theories into a learning trajectory framework that emphasizes the joint influence of curricular goals, classroom culture, and individual prior learning on student learning trajectories. We then test the utility of this framework by analyzing student assessment results from a design research curriculum study that allows us to identify the individual effects of curriculum, classroom, and prior learning on student learning trajectories of fractions concepts over the course of an academic year.

Background and Literature Review

Our proposed framework is based in several bodies of literature, the first of which examines learning trajectories generally and their application to mathematics curricula specifically. The first appearance of the idea of a learning trajectory is often credited to Simon's (1995) account of a teacher-researcher (himself) designing and implementing a sequence of lesson activities in order to support his students in extending their thinking towards a learning goal. This seminal work provided the foundation for studies seeking learning trajectories that are productively associated with student learning. Empirical validation of certain learning trajectories has been influential in developing and refining published curricula such as *Connected Mathematics* (Cain, 2002), *Everyday Mathematics* (Carroll, 1997) and *Building Blocks* (Clements, Sarama, Spitler, Lange, & Wolfe, 2011). In aggregate, these studies have indicated that well-constructed curricular goals and lesson sequences are likely effective in promoting students' individual learning. Furthermore, Empson (2011) eloquently argues that learning trajectories are so interrelated to classroom practices and instructional tasks - including curriculum goals and sequences - that they cannot productively be separated. Therefore curriculum learning trajectories are a key factor in our framework (Figure 1).

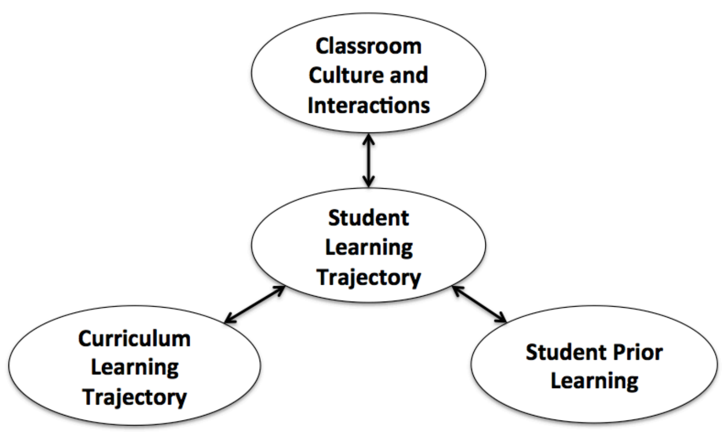


Fig. 1. A learning trajectories framework showing the influences of (a) curriculum learning trajectories, (b) classroom culture and interactions, and (c) student prior learning on students' individual actual learning trajectories.

However, Simon made the point that a learning trajectory was only *hypothetical* until a student actually constructed his or her own new understandings in what is often termed an *actual learning trajectory* (e.g., Battista, 2011). Thus Simon's work elucidated a tension between curricular goals and a student's actual progression through stages of understanding. Other authors have also pointed to this tension (Battista, 2011; Cobb & Bowers, 1999; Lesh & Yoon, 2004), and suggested that differences in individual prior learning and in classroom practices (Empson, 2011) may be some of the factors that contribute to this tension. Therefore we treat the students' actual learning trajectory as the outcome measure in our framework (Fig. 1).

The second body of literature supporting our framework is based in constructivist epistemology and examines individual construction of mathematical understandings. This work indicates that new understandings do not replace prior knowledge but are built through a process of assimilation in which new information is integrated with prior knowledge (Piaget, 1970). Individual differences in prior experiences and understandings may explain why actual learning trajectories sometimes appear non-linear, with movements back and forth or even "off" the hypothesized learning trajectory. Alonzo (2011) even proposed a "landscape of learning" as a more appropriate metaphor because a "landscape" allows teachers and researchers to consider more types of movement than a "progression," which may imply linear movement within a single possible sequence of understandings. Similarly, Battista (2011) suggests using a "profile" to depict an individual student's understandings of a topic in relation to the tasks with which that student engages. Battista suggests that a simple "location" of a learning trajectory gives a very incomplete picture of how the student is making sense of the topic in different contexts (e.g., with different tasks) at a particular time, and a "profile" may be more appropriate for formative assessment purposes. While hypothetical learning trajectories may be useful for guiding instruction, individual students' progression of understandings while engaged with various kinds of instructional tasks may be more complicated, in part due to their prior learning associated with the topics and task types. We therefore include in our framework student prior learning as a factor affecting students' actual learning trajectories (Fig. 1).

Finally, our framework draws from a sociocultural perspective, which suggests that social influences based in the classroom context may also influence individuals' actual learning trajectories. Cobb and Yackel's (1996) *emergent perspective* asserts that social and individual factors are equally important contributions to learning. Murata (2013) builds on these ideas, proposing an instructional model emphasizing the importance of individual students' interactions in the classroom context for their separate, yet related, learning trajectories. Similarly, Saxe and colleagues (2013) analyzed how students and teachers jointly constructed classroom socio-mathematical norms that in turn influenced student learning. These studies, and many others, collectively suggest that classroom culture and interactions are an important influence on how students learn mathematics and therefore likely have a strong impact on actual learning trajectories. This important influence is therefore shown at the top of our framework diagram (Fig. 1).

Thus prior research has shown influences on individuals' actual learning trajectories from curriculum learning trajectories, from individual prior learning, and from classroom culture and interactions. The current study aims to (a) combine these perspectives from prior research into a learning trajectories framework model, and (b) to determine the utility of this framework for understanding curriculum, classroom, and individual influences on students' learning trajectories related to fractions concepts.

Goals of this Investigation

The impetus for this investigation came from two sources: the first was our own interest in the complexity of the learning trajectory construct, which we describe above; and the second was the possibilities inherent in the rich data set that resulted from a recent efficacy study of a two-unit elementary mathematics curriculum. We were fortunate to be part of this efficacy study, and we noted interesting differences in classroom practices (Gearhart et al., 2013) and in individual prior experiences (Saxe et al., 2012), which we believed might contribute to variations in students' actual learning trajectories as they interacted with the curriculum. Additionally, the data included administration of a paper-and-pencil student assessment at four time-points during the school year, making it possible to compare students' responses to particular assessment items over time.

In this investigation, we focused our efforts on identifying the influences of individual, classroom, and curriculum differences on students' learning over time. In operationalizing students' actual learning trajectories, we selected assessment items related to understandings of rational number (i.e., fractions). We identified three research questions, each of which we believe builds off the prior questions:

1. Do individual students exhibit common understandings on a test of related fractions tasks over time? If so, these findings would indicate that students follow similar actual learning trajectories. If not, these findings would suggest that students follow unique learning trajectories, and an investigation to characterize individual differences in trajectories between students (a) interacting with a targeted mathematics curriculum (b) in different classrooms would be appropriate.

2. Do students in the same classrooms exhibit similar learning trajectories over time? If so, these findings would indicate that varying classroom-level factors (e.g., culture, interactions, and practices) may influence how children make sense of fractions concepts. If not, these findings would indicate that students tend to have certain ways of making sense of fractions concepts independent of the influence of classroom culture or practices that may vary across classrooms.

3. Do students' actual learning trajectories align with the learning trajectory of the experimental curriculum? If so, these findings would corroborate prior research that indicates that curricular goals and lesson sequences potentiate student learning of a targeted concept. If not, these findings would provide evidence to suggest that other critical factors outside of the curricular goals significantly interact with the effects of the curriculum to produce different student outcomes.

In the next section, we discuss the data sources and research methods we used to address these three questions. Then, we discuss our findings and the implications of these findings for future research and instructional practice in classrooms.

Methods

Data sources

Our data come from a larger curriculum efficacy study that employed a targeted assessment of fractions and integers concepts and skills administered at multiple time-points. The Learning Mathematics through Representations (LMR) curriculum project is a multi-level design research project that has taken place over the last several years. Initial interview and tutorial studies (Saxe et al., 2010; Saxe, Shaughnessy, Gearhart, & Haldar, 2013) explored elementary students' understandings of integers, fractions, and conventions of the number line, and informed the development of nineteen lessons utilizing the number line to help students coordinate key understandings of integers and fractions. These lessons were individually tested in classrooms and refined with teacher input, and ultimately implemented in eleven fourth- and fifth-grade classrooms in an efficacy study during the 2010-2011 academic year (Saxe et al., 2012).

To evaluate efficacy, a targeted LMR assessment (Saxe et al., 2012) was administered to students at four time points: (a) pre-test before students were exposed to the LMR curriculum, (b) an interim test after the Integers unit of the LMR curriculum was implemented, (c) a post-test after the Fractions unit was implemented, and (d) a final test at the end of the school year. The assessment was also administered at the pre-test, post-test and final time-points to students in ten comparison classrooms who did not receive the LMR curriculum but whose classrooms used the same main curriculum as the students in LMR classrooms - *Everyday Mathematics*. The LMR assessment contained a subset of items that were identical across all administrations of the test, while the rest of the items were adjusted to increase the overall difficulty of the test as students' ability increased. The test incorporated items from the LMR curriculum as well as *Everyday Mathematics* and NAEP released test items.

We limited our investigation to a hypothetical learning trajectory related specifically to the progressive coordination of subunit intervals and unit or multiunit intervals on the number line (as related to understandings of fractions), and we identified the "waypoints" (or levels) of this trajectory through a qualitative review of curriculum materials and interviews with curriculum authors. We selected a subset of five repeated items from the targeted assessment that: (a) involved fractions, (b) were free response items that would reveal students' mathematical ideas or were multiple-choice items wherein distractors were designed to reveal students' mathematical thinking, and (c) were answered incorrectly by the majority of students at pretest and therefore provided an opportunity to reveal conceptual and procedural development (though we do not discriminate between the two; c.f., Schneider, Rittle-Johnson, & Star, 2011) over the course of subsequent tests. (See Appendix for images of selected assessment items.)

Coding of student responses

We used quantitative methods to analyze student responses to the subset of assessment items over time. Using an item response modeling approach (Wilson, 2009) we developed and piloted a coding scheme (Table 1) to categorize specific responses to individual assessment items as indicative of particular levels along the hypothesized learning trajectory from the curriculum.

We are aware of criticisms of the practice of using IRT models to create and/or evaluate learning trajectories (e.g., Battista, 2011). We believe we have addressed these criticisms by using methods that combine IRT with qualitative approaches. For example, Battista (2011) criticizes the use of IRT methods because these models assume that right or wrong answers to a particular assessment item are indicators of the student being "at" a particular level - a level that the student is assumed to be "at" across all assessment items. We address this issue by using open-ended response items and multiple-choice items that were coded not merely as correct or incorrect but rather according to the *type* of thinking indicated by the type of response given (see Table 1). Additionally, the curriculum's learning trajectory was not developed using IRT methods. The LMR curriculum instructional goals and sequence and were developed through a series of investigations utilizing student interviews, teacher interviews, tutorial studies, and other methods to develop an understanding of students' common pathways of learning fractions concepts using the number line model.

Table 1. Curriculum Learning Trajectory and Related Coding Scheme
Levels increase in coordination of ideas from the last row of the table upwards to the first row

Level of Understanding of Rational Numbers (as Fractions) in a Number Line Representational Context	Expected student coordination of ideas while solving tasks on a number line
Response Code A *Coordination of subunit-unit relationships*	Student is likely coordinating (in addition to below): · How to identify the numerator and denominator values using the number line representation, even when the unit interval is not marked · the conceptual meaning of the numerator · that the numerator can be greater than or less than the denominator, and what these two different conditions mean about the value of the fraction

Response Code B *Confusion about subunit counting for numerator only*	Student is likely coordinating (in addition to below): · How to identify the denominator of a fraction · To count some subunits to find the numerator Student is not yet coordinating: · the conceptual meaning of the numerator (how to apply this principle flexibly to identify the numerator value even when the number line context changes; for example, when the unit interval from 0 to 1 is not marked) · that the numerator can be greater than or less than the denominator, and what these two different conditions mean about the value of the fraction
Response Code C *Confusion about subunit counting for denominator*	Student is likely coordinating (in addition to below): · To count some subunits to find the numerator and denominator Student is not yet coordinating: · the conceptual meaning of the numerator and denominator (how to apply these principles flexibly to identify the numerator and denominator values of a fraction, even when the number line context is slightly altered; for example, when the unit interval [from 0 to 1] is not marked) · that the numerator can be greater than or less than the denominator, and what these two different conditions mean about the value of the fraction
Response Code D Initial Fraction Ideas	Student is likely coordinating (in addition to below): · that fractions are values in-between whole numbers · the basic structure of fraction notation (number on top, then line, then number on the bottom) Student is not yet coordinating: · what values the numerator and denominator of a fraction represent

	· how to connect given information on the number line (in terms of unit intervals and subunit intervals) to the meaning of the numerator and denominator
Response Code E *Whole Number Only*	Student is likely coordinating: · the order of counting numbers (whole numbers) · how to identify whole numbers given on the number line · that intervals of the same value must be the same length on the number line Student is not yet coordinating: · that locations in between whole number values on the number line must be fractions · how to identify a given fraction value on the number line

Profile creation

To characterize a student's overall level of coordination of subunit-unit relationships at each assessment time point, we utilized a latent class cluster analysis to condense the five response codes at each time point (one per item) into a single profile or category (Jeroen K. Vermunt & Magidson, 2002). In this analysis, each of the five response codes were entered into the model as dependent variables, with a single latent variable defining the classes. Because we wanted to test for changes over time, we did not impose interdependence between time points and thus each set of student responses was treated as independent by the model. Using the Bayesian Information Criterion, which accounts for parsimony as well as log likelihood of model fit, we selected a 6-class model as the best fit. Therefore, for subsequent analyses each student's outcome variable is a single profile that takes the value 1 (least coordinated) to 6 (most coordinated) at each time point.

It is worth noting that it was very rare for a student to respond with exactly the same response code across all five items at any given time point, unless they were correct on all five, which happened with greater prevalence at the later time points. We believe this to be at least partly the result of a difference in difficulty level between the five selected items, such that a student may find it more accessible to coordinate multiple fraction concepts on an easier problem than on a more difficult one. This was borne out by the initial IRT analysis showing that these five items had varying difficulty levels, and also in the cluster profiles. Students sorted into the first (lowest) cluster profile often got the easiest item correct, while students sorted into the higher profiles often got the more difficult items incorrect. However, to be sorted into the highest profile a student needed to have selected a response code reflecting greater coordination on the items he or she got incorrect. Thus, although each profile included variable response codes across items, that variability was relatively small, and reasonable given different levels of difficulty among assessment items.

Analysis plan

Each of our research questions addresses the influence of one factor from our framework on a student's actual learning trajectory: student prior learning, classroom culture and curriculum learning trajectory. To answer these questions we conducted summary descriptions of the range of student profiles over time as well as several tests of variance within and between treatment groups and classrooms. However, because these methods can illuminate the differences between time points, but not the transition between them, we also conducted a Latent Markov analysis, which shows the probabilities associated with transition between profiles over time (Vermunt, Langeheine, & Böckenholt, 1999). Because this analysis does take into account within-student dependence over time, the latent class profile analysis was redone and the resulting best-fit model had four classes instead of six. Students' two profile assignments were highly correlated to each other at each time point ($r=.83 - .91$, $p<.0001$).

Results

Initial analyses from the efficacy study revealed a significant difference in performance on the targeted LMR assessment between students who received the research curriculum and those who did not (Saxe et al., 2012), providing evidence for the expected positive effect of curricular goals on student learning. We now describe the results of this investigation corresponding to each of our three research questions. First we describe the distribution of individual students' responses and profiles. Next we assess the fit of the individual LTs' alignment to the curriculum learning trajectory of the experimental curriculum. Finally we test for effect of classroom on individuals' actual LTs.

Patterns in individual student understandings

As described above, we take the students' coded responses to be demonstrations of their thinking on each assessment item in our five-item subtest, and the profile (1-6) as the aggregated level of understanding across those items. To determine the similarity in students over time we calculated the percentage of students who were sorted into each profile based on their response codes (Table 2). We found that each profile from least to most sophisticated was represented at each time point, but with varying prevalence. Students overall showed growth over time, with the lower profiles diminishing and the higher ones increasing as students learned the concepts associated with rational numbers.

Table 2. Percentages of students classified into profiles at each assessment time point. *Profiles increase in coordination of ideas from left to right*

Profiles	1		2		3		4		5		6	
Times	LMR	Comp	LMR	Comp	LMR	Comp	LMR	Comp	LMR	Comp	LMR	Comp
Pretest	3.8%		49.5%		14.6%		21.0%		2.4%		8.6%	
	3.0	4.5	49.6	49.4	15.4	13.9	22.6	19.5	2.3	2.6	7.1	10.1
Interim	1.9		46.2		12.1		18.9		0.4		20.5	
	1.9	N/A	46.2	N/A	12.1	N/A	18.9	N/A	0.4	N/A	20.5	N/A
Posttest	1.3		36.6		12.2		11.4		3.4		35.1	
	0.8	1.9	18.8	54.2	12.6	11.7	6.1	16.7	4.2	2.7	57.5	12.9
End of	1.1		21.7		11.3		15.6		6.4		43.9	
year	1.1	1.1	17.5	25.9	8.6	14.1	12.3	19.0	4.5	8.4	56.0	31.6

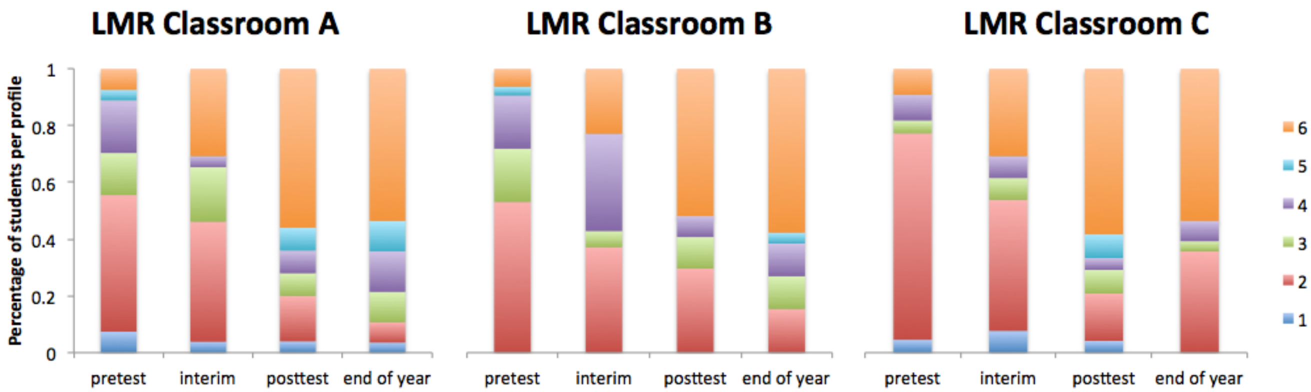
These results suggest that students' actual learning trajectories varied individually, even within groups of students who interacted with the same research-based curriculum. We therefore also investigated variation in response profiles at the classroom level to see if the differences in student responses might be connected to classroom-level effects.

Patterns in item responses and variation between classrooms

We further explored the response patterns of students in the LMR group to our five-item subtest by investigating whether different classrooms showed different response patterns over time, which would suggest classroom-level effects on students' actual learning trajectories.

The curriculum learning trajectory (Table 1) predicts that individuals would progress from lower to higher levels of coordination of ideas (i.e., in the direction of profile 1 to profile 6) over the course of LMR instruction. While this general prediction appears to be true for many classes, as indicated by the increase in profile 6 over time (Table 2, Figure 2), the frequency of each profile varies by classroom at each time point except at pretest, where the distribution is fairly similar. For example, Classroom A has students at each profile at the end of the year, although most are in the highest profile, 6. Figure 2 illustrates some interesting classroom differences as well. In Classroom B, profile 4 was much more popular at the interim assessment than in any other classroom, and Classroom C's students had more students that did progress beyond profile 2, a lower level of understanding, than in the other classrooms. However, these effects were not statistically significant in a test of variance within the LMR group.

Fig. 2. Each chart depicts one LMR classroom's distribution of profiles over 4 testing sessions. Profiles indicate greater (6) to lesser (1) levels of coordination of fraction concepts (unit, multiunit, and subunit on the number line). As a whole, these results display both similarities across classrooms (e.g., higher proportion of correct responses over time) and substantial differences (e.g., profile 4 is popular in classroom B but rare in classroom C).



In a separate analysis of classroom effects within the Comparison group, we did find evidence of classroom environment affecting student LTs. A generalizability study showed that the effect of teacher contributed a significant amount of variance to students' profiles (44%) when compared to time (48%). A hierarchical linear regression supported that result (see Table 3).

Table 3. Hierarchical linear regression of classroom effects in Comparison classrooms
Significant effects in bold font

	Estimate	Standard Error
Fixed Effect		
Time	**.50**	.04
Random Effects		
Classroom	**.35**	.09
Classroom x Time	**.15**	.03
Student	.79	.11
Student x Time	**.13**	.07
Residual	**1.1**	.03

Alignment of actual LTs to curriculum LT.

As mentioned above, an initial efficacy study demonstrated a significant difference between LMR and Comparison students on the targeted assessment (Saxe et al., 2012). We replicated this finding in the analyses we conducted using a subset of assessment items that were common across time points. We found that there was a higher frequency of correct responses (1-5) and more students who were classified in the upper profile levels (1-6) in the LMR group than the comparison group at posttest (t_{575} = 10.93, p <.0001; t_{523} = 12.00, p <.0001) and end of year assessments (t_{575} = 4.08, p <.0001; t_{529} = 4.74, p <.0001). However, these statistics are based on group aggregates at each time point. To analyze movement between profiles, we utilized a Latent Markov Model.

Latent Markov modeling is similar to latent cluster analysis in that it sorts individual cases into latent classes based on outcome measures (here, five response codes), and then calculates the transitional probability than an individual in each class will move to each of the other class profiles at the next time point. It also takes into account dependence within individuals across time points in constructing the latent classes, so this analysis resulted in a best-fit model with four profiles instead of six. Because our Comparison group did not take the interim test, we utilized only the response codes from pretest, posttest, and end of year assessment for parity. Typically, the time points do not matter in Latent Markov modeling as the process is assumed to be dependent only on the current state. However, because student learning was at least in part dependent on the curriculum, which varied between time points, it was important that we base our analyses on the unique transitional probabilities for each time point.

Although our results were not significant, likely due to the high number of parameters (multiple time points by two groups) in our model, they do indicate that the students progress upwards through the learning trajectory profiles in a mostly linear fashion (Table 4). From pretest to posttest, the LMR students who start in profile 1 are most likely to progress to profile 3, but profile 2 is also common. For LMR students who start in profiles 2, 3 or 4, at the first assessment, they are all most likely to progress to the most advanced profile, 4. If we include data from the LMR group at the interim assessment, we see more incremental, and thus linear, movement through the profiles. Students who advanced to level 4 at post-test tended to maintain these gains (their level of coordination) at the follow-up end of year assessment, with the exception of students in profile 3, who either slip or advance. In contrast, the Comparison group is most likely to remain in the profile they started from pretest to posttest, and then progress incrementally by one profile in the second half of the academic year.

Table 4. Transitional probabilities of moving from one profile to another
Profile 1 shows the least coordination of fraction ideas and profile 4 shows the most coordination.

LMR — Posttest
Pretest

Profiles	1	2	3	4
1	.17	.28	.49	.05
2	.00	.39	.00	.60
3	.00	.01	.00	.99
4	.00	.01	.00	.99

Comp — Posttest
Pretest

Profiles	1	2	3	4
1	.51	.48	.00	.00
2	.00	.97	.03	.00
3	.00	.00	.85	.15
4	.00	.00	.00	.99

LMR — End of Year
Posttest

Profiles	1	2	3	4
1	.31	.62	.05	.01
2	.03	.57	.36	.04
3	.00	.55	.00	.45
4	.00	.00	.08	.92

Comp — End of Year
Posttest

Profiles	1	2	3	4
1	.46	.54	.00	.00
2	.00	.32	.68	.00
3	.00	.00	.28	.71
4	.03	.00	.01	.96

These preliminary results showing relatively linear progression from less advanced to more advanced levels of understanding indicating that the curriculum LT embedded in LMR reflects a natural learning progression. This is because the Comparison group shows a linear progression even more so than the LMR group. However, due to the fact that these are statistically non-significant, additional research must be conducted to clarify these results.

Conclusions

This LMR dataset affords a rare opportunity to examine individual, classroom, and curriculum influences on learning trajectories over time. By demonstrating that each of these factors impacts individual students' actual learning trajectories, this dataset supports the multifaceted learning trajectory framework we proposed here. Additional research to characterize the mechanisms surrounding the interactions between these factors will provide insight into how to leverage them optimally. This information may be useful not only for education researchers but also for practitioners in that it suggests possible areas for intervention at the level of curriculum goals and sequences, classroom culture and interaction, as well as individual understandings. Furthermore, we hope that greater exploration of these effects separately as well as their interactions will lead to a more precise vocabulary describing the various levels of effects on learning trajectories (e.g., hypothetical learning trajectories vs. actual learning trajectories), and may assist researchers and practitioners in coordinating their efforts around implementation of new curriculum materials.

Addendum

Since presenting the conference proceedings that comprises Chapter 3 of this dissertation, Dr. Geoff Saxe and I have re-conceptualized the analysis. The construct of learning trajectory has been described at the individual level (Simon, 1995), and within the context of design research in the classroom (Cobb & Bowers, 1999) – two approaches that describe how instructors adapt their learning activities based on the demonstrated understandings of their students. Battista and Confrey have both emphasized the use of learning-trajectory-based assessments to assist instructors in diagnosing their students' understandings for the purpose of adapting instruction (Battista, 2011; Sztajn, Confrey, Wilson, & Edgington, 2012). However, to my knowledge only one researcher has investigated the range of differences in longitudinal learning trajectories between individuals (Wright, 2011). The variability recorded in that analysis, as well as the initial findings from our own conference proceedings, suggests a scalability issue for teachers. Specifically, with different students in a classroom demonstrating different levels of understanding at any given time, teachers may be required to flexibly restructure learning activities in response to a wide array of student thinking. It is unclear how teachers handle this range of variability in their classrooms, which, given recent standards regarding conceptual development, is of utmost importance in ensuring all students have adequate opportunity to learn.

The follow-up investigation based on Chapter 3 focuses on the variability of students' levels of understanding within different classrooms in the LMR study. My goal is to characterize the extent of variation in student understandings that teachers may face related to a specific learning trajectory. Secondly, I hope to identify some best practices in handling that variability or mitigating it such that all students may engage productively in their learning environments.

Closing Remarks

Summary of Findings

In Chapters 1 and 2 of this dissertation, I investigated the cognitive processes involved in the performance of a proportional reasoning task, interpreted the findings regarding the underlying mental constructs supporting proportional thinking, and suggested some implications for pedagogical settings. Within the chapters, I interpreted the findings only within the scope of the theories established in the relevant literature. Here I will tie the findings back to the theories outlined in the Introduction of this dissertation to draw a broader conclusion.

The developmental comparison described in Chapter 1 drew a clear distinction in cognitive processes between children and adults. The children with comparable accuracy to the adults' performance on the fraction comparison task nevertheless demonstrated eye movements akin to their low-performing peers, and quite different from the adults. This led to the inference that the cognitive processes elicited by this task were influenced to a greater extent by development and experience than by mathematical performance. However, an important similarity between the high-performing children and adults is that they both exhibited slower response times and relatively more eye movements in the more difficult conditions, although the effect was weaker for children. In contrast, the low-performing children answered just as, if not more, quickly when the task was difficult, indicating either a lack of recognition of the task demands or a lack of persistence. Thus, greater task proficiency was associated with more adaptive allocation of attention.

Revisiting these findings through the lens of relational complexity theory, it is necessary to first evaluate the complexity of the task conditions. The two easier task conditions had a binary effective complexity: although there were four numbers presented, either the denominators (SD condition) or the numerators (SN condition) were identical, so only a binary comparison between the non-identical components was necessary. The two more difficult conditions required attention to all four numbers, which could elicit either quaternary comparisons, or a series of binary comparisons, as described in the Introduction.

Based on relational complexity theory, one would predict that the effective complexity of the given pair has the greatest impact on performance. If that were the case, the first two conditions – same denominator (SD) and same numerator (SN) – should be behaviorally identical. However, our data showed a distinction between these two conditions. Although the highly-proficient groups performed with high accuracy on both conditions, the adults responded more quickly to the SD than SN problems. This difference indicates the presence of an influential factor beyond relational complexity; a likely candidate is the general cognitive capacity of inhibition, which influences performance in a variety of domains (e.g., Davidson, Amso, Anderson, & Diamond, 2006). Specifically, the SN condition requires inhibition to override the prepotent response of selecting the larger number when that number is in the denominator position. Indeed, Meert, Grégoire, and Noël have already offered empirical evidence of an inhibitory effect in children's performance on this task (Meert et al., 2010b), although they did not test this effect in adults.

Thus, the additional cognitive load required to inhibit a prepotent response could explain the decrement to performance we see in adults for SN compared to SD trials.

The additional cognitive load imposed by SN over SD trials dovetails with educational theory that posits denominators are a more advanced concept to construct, which is a second potential influence beyond relational complexity. The finding that adults demonstrate this disadvantage when handling more advanced concepts suggests that, despite years of experience with fractions, they have not automatized the conceptual thinking required to complete the fraction comparison task. It is possible that the experimental conditions, the unfamiliarity of the task, and the fact that this task was the first in the session, made them more cautious and enhanced the effect of the conceptual difficulty related to denominators. It would be necessary to manipulate the order of tasks within the session to determine whether conceptual difficulty or cognitive load due to inhibition was more impactful to adults' performance.

It is important to note that we saw this SN-related performance decrement only in adults, and not in children. Both children's relative cognitive immaturity and their status as novice fraction reasoners would lead one to predict that the effects of inhibition and conceptual difficulty would be more, not less, pronounced than adults. Instead, we saw no evidence that children responded to SN differently than SD, except in the types of eye movements they made. An additional potential factor could explain this effect: cognitive flexibility, another general cognitive skill that is still immature in children of this age (Davidson et al., 2006). The task required participants to select the larger fraction, which results in opposing rules for the SD and SN conditions: select the larger number in the numerator position when denominators are equal, and the smaller number in the denominator position when numerators are equal. Because the trials were pseudo-randomly presented, children had to attend closely to each trial and select the appropriate rule, which would impact SN trials as much as any other condition. Moreover, because this group was just learning fractions, they may have found even the relatively easy SD trials to be conceptually challenging; if that were the case there would be no benefit of SD over SN. To determine the relative effects of these two potential factors, we would need to compare performance on this interleaved task to a blocked version of the task in which SD and SN trials were presented separately. If the blocked design were to result in a performance difference between SD and SN, we may conclude that the lack of difference shown here is due to the cognitive load imposed by flexibility counteracting the conceptual facility with numerators over denominators.

Although the above findings do not support predictions related to relational complexity theory, there are some indicators that do. Specifically, the IC condition, in which participants necessarily had to attend to all four numbers, was associated with poorer accuracy and slower response times in both high-proficiency adults and children. English and Halford (1995) suggested that quaternary comparisons, which participants could engage in on CO and IC trials, could be resolved by one of two approaches: the unit rate and factor-of-change approaches – which, as explained in the Introduction, are akin to structure-mapping and project-first theories of analogy solving. In addition to these approaches, more experienced mathematicians may have found it easiest to estimate the relative magnitude of each fraction, or to execute a cross-multiplication strategy. Our analyses do not provide insight as to which of these approaches participants used, but the IC-related performance decrements suggest that one of these more complex solution strategies was likely used.

The CO condition is difficult to conceptualize with regard to relational complexity theory, because the trials were arranged such that binary comparisons between numerators or between denominators would lead a participant to select the same answer. Thus, although CO trials appear to be quaternary comparisons, they could effectively be solved with only binary comparisons. The effective complexity may be the reason CO trials show more similar performance to SD and SN than to IC trials. If this were the case, it would enhance support for relational complexity theory, which would lead one to predict a clear performance distinction between the effective levels of complexity.

The fact that the low-performing children did not modify their behavior across the different task conditions suggests they were not sensitive to either the increased relational complexity or the conceptual challenges associated with denominators relative to numerators. However, an alternate explanation is that they were sensitive to these factors but didn't have access to any of the more advanced strategies and thus chose to guess quickly at random. These general findings align with both relational complexity theory and educational theorists' description of conceptual construction. Combined with the other findings noted above, we can see that theories from cognitive science and education all provide fitting explanations for different aspects of these data, gleaned from a relatively simple experimental task.

However, the unexplained factors in the adult performance led me to question which cognitive strategies were most helpful to experienced proportional reasoners, and in Chapter 2 I described the results of an experiment analyzing adult performance on a much more difficult version of the fraction comparison task. In that experiment, all of the pairs represented quaternary relations, and so relational complexity theory did not apply. Instead, I interpreted the results in terms of analogical reasoning theories, describing the eye movements as within-fraction or between-fraction. The primary finding was that people do adjust their strategies based on the affordances of the task at hand. This finding supports the mathematical theory that people hold a hybrid mental representation of fractions, in that they can access both magnitudes and individual fraction components as needed, and also extends the literature. The hybrid model has been previously supported by other studies, but only in analyses of fraction pairs that shared common components, e.g., 2/7 vs 5/7, contrasted with pairs that do not share components, e.g., 3/5 vs 4/7. In the common-components case, only a simple first-order comparison between the two non-identical components is necessary, and therefore componential processing is to be expected. The results of the study reported in Chapter 2 go well beyond those initial findings, in that our study showed evidence of componential processing in the context of fraction pairs that had no shared components. Additionally, prior research had only reported evidence for holistic magnitude processing in that context.

Because prior research supported the benefits of magnitude processing in the case of no shared components, we predicted that a higher proportion of within-fraction comparisons would lead to greater overall success, and particularly to better performance on the task conditions that favored a magnitude-focused strategy. However, our data did not support that hypothesis. Instead, it seems that between-fraction comparisons are beneficial in specific task conditions – in this case, when one denominator is an easily-recognizable multiple of the other, which facilitates a conversion strategy – but not in general. Moreover, we found initial evidence that people who more consistently used that strategy demonstrated faster overall response times. Further, we did not find a benefit to focusing on magnitudes, even when the task condition promoted magnitude processing.

Taken together, and in light of the theories presented in the Introduction, these studies support the claims that relational complexity may exert a partial influence on reasoners' performance, but does not completely explain behavior on this proportional reasoning task, and theories from both education and cognitive science offer additional explanatory value. In particular, further research testing more nuanced conditions will be necessary to characterize adults' mathematical reasoning. Adult behavior on the more difficult, adult-oriented task did reflect structure-mapping theory, in that the relations between fractions were attended to a greater extent than the relations within fractions. However, this contradicts previous research (Obersteiner & Tumpek, 2016) and therefore should be a particular focus for future research to better characterize adult strategies.

Ultimately, the reason for characterizing mature proportional reasoning is to support the creation of instructional materials that emphasize concepts, strategies, and problem-solving approaches that will best serve students. Of course, much is already known about how children learn proportional reasoning concepts, as is reflected in Chapter 3, and my purpose for including both approaches in this dissertation is to demonstrate the breadth of insight that can be gleaned by attending to multiple disciplines. In particular, Chapter 3 showed the impact that social and environmental factors have on learning about rational numbers, in addition to the cognitive factors described in the eye-tracking studies. Students' learning trajectories over the course of an academic year are influenced by their prior knowledge of rational numbers, as evidenced by the understandings they demonstrate at the beginning of the year; by the classrooms and social practices that situate their learning; and by the curriculum with which they engage, as evidenced by the dramatic differences in learning trajectories between the LMR and Comparison groups. Thus, attending to the social and environmental factors that impact learning is critical to effectively incorporating learnings from cognitive psychological research into pedagogical practice.

Extensions of current research findings

This dissertation is based on the claim that proportional reasoning and relational reasoning share cognitive foundations. In this section, I first critique that thesis as a means of suggesting additional research in relational and proportional reasoning, separately. I have also, throughout this dissertation, emphasized my belief in the importance of attending to both cognitive and educational research traditions, and noted above that theories from several different fields explain aspects of the studies I presented in Chapters 1 and 2. Chapter 3 describes research undertaken in a classroom setting, and in so doing it underscores the complexity of real-life learning, which only increases the challenge of invoking appropriate explanations for research observations. In this section I would like to extend these ideas a bit farther and speculate on how these studies might be interpreted in light of one another. Although doing so necessarily confuses the picture, I assert that it is an important step in identifying potential areas for future cognitive psychology and educational research.

My initial motivation for applying analogical reasoning theories to proportional reasoning is because children spontaneously use analogical reasoning skills, which are generally well developed by the time fractions are typically introduced in schools. Therefore, if the cognitive mechanisms that support analogical reasoning could fruitfully be applied to mathematical reasoning, they may serve as useful pedagogical tools regarding proportions. The study of analogical reasoning encompasses a range of cognitive mechanisms, such as encoding relational structures, mapping analogs, and applying relations – all of which may be important in proportional reasoning. Yet the fraction comparison task additionally requires flexibility to identify and utilize different types of relationships within and between the presented proportions. Studies of relational reasoning often address the integration of multiple sets of relations (Crone et al., 2009; Wendelken & Bunge, 2010; Wendelken, O'Hare, Whitaker, Ferrer, & Bunge, 2011), or the ability to notice and map relations between complex analogs (Gentner, 1983; Gick & Holyoak, 1987), but not necessarily their identification or selection when multiple sets of relations between and within item pairs are equally valid.

I highlight this disconnect as support for the following suggestions. First, it may be helpful to better characterize the role of cognitive flexibility regarding the identification and selection of relationships within a non-mathematical reasoning context. Doing so may further our understanding of the cognitive foundations of relational reasoning broadly defined, the factors that promote or hinder it, and the strategies that are associated with proficiency in this general domain. Second, although this dissertation is based on the thesis that relational reasoning and proportional reasoning share essential cognitive foundations, a key difference between these domains is the fact that all numbers are related to each other along a common scale of magnitude. This common scale affords the flexibility to use either within-fraction or between-fraction relations when solving proportional reasoning problems, in a way that classical analogies may or may not support. The differences between the structure-mapping and project-first theories of analogical reasoning highlight within- versus between-analog approaches, and yet I only know of a few studies that systematically test which analogical reasoning theory is better supported empirically (Thibaut et al., 2011; Vendetti, Starr, Johnson, Modavi, & Bunge, in revision). The lack of empirical resolution may be because most tasks lend themselves to one method or the other and thus both theories are valid in certain contexts, or because the difference between them is behaviorally indiscernible. Eye-tracking methodology provides an opportunity to distinguish between these approaches, given an appropriate analogical task, and thus may be a useful means for investigating flexibility in a non-mathematical relational reasoning domain.

However, the common scale of numbers and the resulting flexibility in relations may be the feature that definitively separates proportional reasoning from analogical reasoning, on a cognitive basis, and perhaps the study of proportional reasoning should focus not on the overlaps but on the distinctions between these domains. Even if this were the case, I maintain that it is important to characterize proficient strategies in a proportional reasoning context. While a person's choice of which relations to attend to may be driven largely by the strategy they learned most comfortably when young, it remains a strong possibility that the ability to 'see' multiple relations and select the most appropriate ones for a given situation is the optimal skill, similar to descriptions of expertise in other domains of knowledge (Feltovich, Prietula, & Ericsson, 2006).

Next, I explore further interpretations of my cognitive studies in light of my educational findings, and vice versa. First, the eye-tracking studies did not address the curricula used by the various participants and students. While undertaking the study described in Chapter 1, I reviewed the curriculum used by the children, which contained only an introduction to fractions and their meanings in rote, traditional senses. There was no comparison of fractions, with the exception of later units that prescribed steps for converting one fraction into an equivalent fraction in order to compare two fractions with the same denominator. Nowhere was there a mention of relative magnitudes or proportional relationships, so the task the students completed on the eye-tracker was very much out of their realm of expertise, and would have been even at the end of the academic year. Had the children been more familiar with the magnitudes of the fractions represented, they may have approached the task quite differently. For example, if they had been instructed using a number line that makes relative magnitudes more apparent, or if they had previously done any magnitude comparisons with fractions, we may have seen less disparity between the groups of children, or better overall performance with the more difficult conditions. As the curriculum stands, their attraction to viewing the relevant numbers may have been an effect of perceptual salience; knowing that their formal education did not include tasks like the one studied means this and other non-mathematical interpretations may be valid.

I also did not collect information about the fraction educations of the adult participants, who comprised the participant sample in both Chapters 1 and 2. Given the findings of Chapter 3, that curriculum does affect people's knowledge and learning of fractions, having this information may have clarified the cognitive findings. I tested primarily for two strategic approaches: comparisons within fractions and between fractions. I also tested for but did not find evidence of the systematic use of a cross-multiplication strategy. However, there may be other algorithms or strategies I could have tested for had I gathered data on the full range of strategies they used. I will note that I did ask participants to self-report their strategies on a variety of problems at the end of the session, and on the whole I did not find their explanations to vary greatly from the strategies I tested for, but a more thorough interview study may have uncovered additional possibilities. It is worth conducting such a study because the goal of this line of research is to characterize the types of cognitive strategies that are beneficial to proportional reasoning and those that hinder it, in order to better inform instructional materials and methods. Both self-reports and objective eye-tracking methods are worthwhile in this pursuit, as people are not always good at articulating their own thought patterns.

Applying a cognitive psychology lens to my educational research, the assessment items that defined the learning trajectories in Chapter 3 reflected a range of mathematical representations, emphasizing the number line but including area models and symbols as well. Indeed, the range of mathematical tasks in curricular materials is generally much wider than that used in laboratory tasks, and for good reason – children need to learn to recognize different applications of fractions, and reason mathematically in different contexts. However, we know from pedagogical analyses of these representations that, although all are represented by the same formal *a/b* notation, each has a different meaning, e.g., parts of a whole, parts of a set, rates, ratios, etc. (Behr et al., 1983). Some of these representations are known to be more challenging than others for young learners (Lamon, 2012), but I am not aware of research that systematically investigates whether these different meanings are associated with different mental representations, or how different perceptual representations may be mapped to the formal notation. Conducting this learning trajectory research in a way that addresses the potential for different underlying cognitive mechanisms may further illuminate how students develop fraction concepts.

In this section I have commented on some issues related to the studies reported in this dissertation. In the section that follows I discuss several separate research endeavors that may tie together on a broader scale the fields of cognitive psychology and education research.

Future Directions

There are several promising lines of inquiry that may further the effort to build an interdisciplinary vocabulary and jointly-relevant research methodology spanning cognitive neuroscience and education. The first is to more explicitly study cognitive processes within educational settings. Richland and colleagues have started down this path with their research that investigates how teaching practices adhere to or could better support analogical reasoning within mathematics classrooms. For example, visually aligning two given examples, verbally comparing them, and using gesture are instructional techniques known to help students draw conceptual connections. Richland et al. found that instructional videos that used all of these cognitive supports led to better performance on a post-test, particularly for a dissimilar problem type, than videos that did not provide these cognitive supports (Richland, 2003). More research like this will clarify which cognitive theories translate to classroom practice, and to what extent they promote learning in complex environments.

Another potentially fruitful arena could grow from the emerging evidence for a proportional percept, similar to the approximate number sense already well-documented in the cognitive literature. Developmental work such as demonstrations of infants' proportional reasoning (e.g., Denison & Xu, 2014), as well as empirical support for Lewis, Hubbard, and Matthew's proposed ratio processing system (RPS; 2015) emphasize the strong intuitions people have about proportions. Yet these early abilities must get dissociated from mathematical reasoning in the course of early education, or else we would not have anecdotes like the 1/3 hamburger problem. An early successful example of building instruction on intuition comes from Moss & Case's experimental curriculum in which they taught percentages before fractions, contrary to traditional methods, and found that children had better intuitions about percentages, and were also better able to connect those intuitions to fractions when taught in a modified sequence (Moss & Case, 1999). More research on the RPS may provide insights for educators on how to take advantage of learners' intuitive proportional knowledge and build on it through instruction.

The second approach to spanning disciplines is to craft lab-based experiments that investigate particular educational phenomena of interest to educators. Although, as stated in the Introduction, much of the current educational research is necessarily embedded within social environments, and Vygotsky stated that these complex social-cognitive systems are impossible to study with experimental methods (1978), there are newer laboratory methods that invoke social influences. Using partner- or group-based experimental designs may produce a lab-based facsimile of the learning environment in order to study specific, reduced research questions using experimental methods. Furthermore, successful collaborations between researchers and educators feature teacher-generated questions about their students' conceptual understanding, that the researchers can design targeted metrics or interventions to analyze. Fostering academic-practitioner collaborations such as these will support the application of research to practice, and use practice to guide research.

Conclusion

The research reported in this dissertation represents my efforts to embody and practice research from both cognitive and educational perspectives. While the conclusions of these particular studies advance, at best, only the fields in which they are currently situated, the experience of conducting these studies has better positioned me to foster collaborations across the disciplines, which I believe to be the best way to promote future advances in both education and cognitive science.

References

Alexander, P. A., & Buehl, M. M. (2004). Seeing the Possibilities: Constructing and validating measures of mathematical and analogical reasoning for young children. In L. D. English (Ed.), *Mathematical and analogical reasoning of young learners* (pp. 23–45). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Alexander, P. A., White, C. S., & Daugherty, M. (1997). Analogical Reasoning and Early Mathematics Learning. In L. D. English (Ed.), *Mathematical Reasoning: Analogies, Metaphors, and Images* (pp. 117–147). Routledge.

Alonzo, A. C. (2011). Learning Progressions That Support Formative Assessment Practices. *Measurement: Interdisciplinary Research and Perspectives*, *9*(2–3), 124–129. http://doi.org/10.1080/15366367.2011.599629

Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. *Nature Reviews. Neuroscience*, *9*(4), 278–91. http://doi.org/10.1038/nrn2334

Awh, E., Armstrong, K. M., & Moore, T. (2006). Visual and oculomotor selection: Links, causes and implications for spatial attention. *Trends in Cognitive Sciences*, *10*(3), 124–130. http://doi.org/10.1016/j.tics.2006.01.001

Baroody, A. J., Cibulskis, M., Lai, M.-L., & Li, X. (2004). Comments on the Use of Learning Trajectories in Curriculum Development and Research. *Mathematical Thinking and Learning*, *6*(2), 227–260.

Battista, M. T. (2011). Conceptualizations and issues related to learning progressions, learning trajectories, and levels of sophistication. *The Mathematics Enthusiast*, *8*(3), 507–570.

Behr, M. J., Lesh, R., Post, T. R., & Silver, E. A. (1983). Rational Number Concepts. In R. Lesh & M. Landau (Eds.), *Acquisition of Mathematics Concepts and Processes* (pp. 91–125). New York, NY: Academic Press Inc.

Behr, M. J., Wachsmuth, I., Post, T. R., & Lesh, R. (1984). Order and Equivalence of Rational Numbers: A Clinical Teaching Experiment. *Journal for Research in Mathematics Education*, *15*(5), 323–341.

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. http://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. http://doi.org/10.1016/j.cedpsych.2012.07.001

Booth, J. L., Newton, K. J., & Twiss-Garrity, L. K. (2014). The impact of fraction magnitude knowledge on algebra performance and learning. *Journal of Experimental Child Psychology*, *118*(1), 110–118. http://doi.org/10.1016/j.jecp.2013.09.001

Boyer, C. B., & Merzbach, U. C. (2011). *A History of Mathematics* (Third). Hoboken, NJ: John Wiley & Sons.

Boyer, T. W., Levine, S. C., & Huttenlocher, J. (2008). Development of Proportional Reasoning: Where Young Children Go Wrong. *Developmental Psychology*, *44*(5),

1478–1490. http://doi.org/10.1037/a0013110.Development
Brannon, E. M. (2006). The representation of numerical magnitude. *Current Opinion in Neurobiology*, *16*(2), 222–9. http://doi.org/10.1016/j.conb.2006.03.002
Bright, G. W., Behr, M. J., Post, T. R., & Wachsmuth, I. (1988). Identifying Fractions on Number Lines. *Journal for Research in Mathematics Education*, *19*(3), 215–232.
Bucci, M. P., & Seassau, M. (2012). Saccadic eye movements in children: A developmental study. *Experimental Brain Research*, *222*(1–2), 21–30. http://doi.org/10.1007/s00221-012-3192-7
Bunge, S. A., & Wright, S. B. (2007). Neurodevelopmental changes in working memory and cognitive control. *Current Opinion in Neurobiology*, *17*, 243–250. http://doi.org/10.1016/j.conb.2007.02.005
Cain, J. S. (2002). An Evaluation of the Connected Mathematics Project. *The Journal of Educational Research*, *95*(November), 224–233. http://doi.org/10.1080/00220670209596595
Calvo, M. G. (2001). Working memory and inferences: Evidence from eye fixations during reading. *Memory (Hove, England)*, *9*(4), 365–381. http://doi.org/10.1080/09658210143000083
Carpenter, T. P., Corbitt, M. K., Kepner, H. S., Lindquist, M. M., & Reys, R. (1980). Results of the second NAEP mathematics assessment: Secondary school. *The Mathematics Teacher*, *73*(5), 329–338. http://doi.org/http://www.jstor.org/stable/27963780
Carroll, W. M. (1997). Results of Third-Grade Students in a Reform Curriculum on the Illinois State Mathematics Test. *Journal for Research in Mathematics Education*, *28*(2), 237–242. http://doi.org/10.2307/749763
Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational Efficiency and the Growth of Short-Term Memory Span. *Journal of Experimental Child Psychology*, *33*, 386–404.
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. http://doi.org/10.1007/s10649-009-9198-9
Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics Learned by Young Children in an Intervention Based on Learning Trajectories : A Large-Scale Cluster Randomized Trial. *Journal for Research in Mathematics Education*, *42*(2), 127–166.
Cobb, P., & Bowers, J. (1999). Cognitive and situated learning perspectives in theory and practice. *Educational Researcher*, *28*(2), 4–15. http://doi.org/10.2307/1177185
Cobb, P., & Yackel, E. (1996). Constructivist, Emergent, and Sociocultural Perspectives in the Context of Developmental Research. *Educational Psychologist*, *31*(3), 175–190.
Cohen, M. E., & Ross, L. E. (1977). Saccade Latency in Children and Adults: Effects of Warning Interval and Target Eccentricity. *Journal of Experimental Child Psychology*, *23*, 539–549.
Common Core State Standards Initiative. (2010). *Common Core State Standards for*

Mathematics. *Learning*.
Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, a Z., Ollinger, J. M., Drury, H. A., … Van Essen et al., D. C. (1998). A common network of functionnal areas for attention and eye-movements. *Neuron*, *21*(4), 761–773. http://doi.org/10.1016/S0896-6273(00)80593-0
Cramer, K., Post, T., & Currier, S. (1993). Learning and teaching ratio and proportion: Research implications. In D. Owens (Ed.), *Research ideas for the classroom* (pp. 159–178). New York, NY: Macmillan.
Crawford, J. L., Miller Singley, A. T., & Bunge, S. A. (n.d.). Using Eye Tracking to Probe Developmental and Skill-based Differences in Fraction Magnitude Evaluation.
Crone, E. A., Wendelken, C., Van Leijenhorst, L., Honomichl, R. D., Christoff, K., & Bunge, S. A. (2009). Neurocognitive development of relational reasoning. *Developmental Science*, *12*(1), 55–66.
Daro, P., Mosher, F. A., & Corcoran, T. (2011). *Learning Trajectories in Mathematics: A Foundation for Standards, Curriculum, Assessment, and Instruction*. *Consortium for Policy Research in Education*.
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. http://doi.org/10.1016/j.neuropsychologia.2006.02.006
Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*(1–2), 1–42. http://doi.org/10.1016/0010-0277(92)90049-N
Dehaene, S. (1997). *The Number Sense: How the mind creates mathematics*. New York, NY: Oxford Univ Press.
Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, *20*(3), 487–506. http://doi.org/10.1080/02643290244000239
Denison, S., & Xu, F. (2014). The origins of probabilistic inference in human infants. *Cognition*, *130*(3), 335–347. http://doi.org/10.1016/j.cognition.2013.12.001
Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*(12), 1827–1837. http://doi.org/10.1016/0042-6989(95)00294-4
DeWolf, M., Grounds, M. a, Bassok, M., & Holyoak, K. J. (2014). Magnitude comparison with different types of rational numbers. *Journal of Experimental Psychology. Human Perception and Performance*, *40*(1), 71–82. http://doi.org/10.1037/a0032916
Dweck, C. (2006). *Mindset: The new psychology of success*. Random House.
Elwes, R. (2013). *Math in 100 Key Breakthroughs*. Quercus.
Empson, S. B. (2011). On the idea of learning trajectories: Promises and pitfalls. *The Mathematics Enthusiast*, *8*(3), 571–596.
English, L. D., & Halford, G. S. (1995). *Mathematics education: models and processes*. Mahwah, New Jersey: 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. http://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. http://doi.org/10.1027/1618-3169/a000116
Fazio, L. K., DeWolf, M., & Siegler, R. S. (2015). Journal of Experimental Psychology : Learning , Memory , and Cognition Comparison Strategy Use and Strategy Choice in Fraction Magnitude Comparison.
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.
Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–14. http://doi.org/10.1016/j.tics.2004.05.002
Feltovich, P. J., Prietula, M. J., & Ericsson, K. A. (2006). Studies of expertise from psychological perspectives. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *Cambridge Handbook of Expertise and Expert Performance* (pp. 41–68). New York, NY: Cambridge University Press.
Fukushima, J., Hatta, T., & Fukushima, K. (2000). Development of voluntary control of saccadic eye movements: Age-related changes in normal children. *Brain and Development*, *22*, 173–180.
Gabriel, F., Coché, F., Szucs, D., Carette, V., Rey, B., & Content, A. (2013). A componential view of children's difficulties in learning fractions. *Frontiers in Psychology*, *4*(October), 1–12. http://doi.org/10.3389/fpsyg.2013.00715
Gallistel, C. R., & Gelman, R. (2005). Mathematical Cognition. In K. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 559–588). Cambridge University Press.
Gearhart, M., Buchanan, N. L., Collett, J., Diakow, R., Kang, B., & Saxe, G. B. (2013). Opportunities to learn mathematics: Role of students' engagement in discussion. In *Research Presession of the Annual Meeting of the National Council of Teachers of Mathematics*. Denver, CO.
Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2013). Adolescents' Functional Numeracy Is Predicted by Their School Entry Number System Knowledge. *PLoS ONE*, *8*(1). http://doi.org/10.1371/journal.pone.0054651
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. http://doi.org/10.1007/s10648-011-9174-7
Geier, C., & Luna, B. (2009). The maturation of incentive processing and cognitive control. *Pharmacology Biochemistry and Behavior*. http://doi.org/10.1016/j.pbb.2009.01.021
Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*(2), 155–170. http://doi.org/10.1016/S0364-0213(83)80009-3
Gentner, D. (1998). Similarity and the development of rules, *65*, 263–297.
Gentner, D., & Christie, S. (2014). Mutual bootstrapping between language and analogical processing. *Language and Cognition*, *2*(2), 261–283. http://doi.org/10.1515/langcog.2010.011

Gick, M. L., & Holyoak, K. J. (1987). The cognitive basis of knowledge transfer. In S. M. Cormier & J. D. Hagman (Eds.), *Transfer of learning: Contemporary research and applications* (pp. 9–46). San Diego, CA: Academic Press Inc.
Goswami, U. (1992). *Analogical reasoning in children*. Sussex, England: Erlbaum.
Grant, E. R., & Spivey, M. J. (2003). Eye Movements and Problem Solving: Guiding Attention Guides Thought. *Psychological Science*, *14*(5), 462–466.
Green, C. T., Mckerracher, A., Whitaker, K. J., Ferrer, E., & Bunge, S. A. (2011). Longitudinal predictors of mathematical reasoning during development. In *Poster session*.
Green, E. (2014, July). Why Do Americans Stink at Math? *The New York Times*, Retrieved from URL.
Grosbras, M. H., Laird, A. R., & Paus, T. (2005). Cortical regions involved in eye movements, shifts of attention, and gaze perception. *Human Brain Mapping*, *25*(1), 140–154. http://doi.org/10.1002/hbm.20145
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. http://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 Paneld*. Chicago, IL.
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 Univ Press.
Huber, S., Moeller, K., & Nuerk, H. C. (2014). Adaptive processing of fractions - Evidence from eye-tracking. *Acta Psychologica*, *148*, 37–48. http://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.
Irving, E. L., Steinbach, M. J., Lillakas, L., Babu, R. J., & Hutchings, N. (2006). Horizontal Saccade Dynamics across the Human Life Span. *Investigative Ophthalmology & Visual Science*, *47*(6), 2478–2484. http://doi.org/10.1167/iovs.05-1311
Irving, E. L., Tajik-Parvinchi, D. J., Lillakas, L., González, E. G., & Steinbach, M. J. (2008). Mixed pro and antisaccade performance in children and adults. *Brain Research*, *1255*, 67–74. http://doi.org/10.1016/j.brainres.2008.12.006
Ischebeck, A., Schocke, M., & Delazer, M. (2009). The processing and representation of fractions within the brain: An fMRI investigation. *NeuroImage*, *47*, 403–413. http://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. http://doi.org/10.1080/17470218.2015.1046464
Jacob, S. N., Vallentin, D., & Nieder, A. (2012). Relating magnitudes: the brain's code for proportions. *Trends in Cognitive Sciences*, *16*(3), 157–166. http://doi.org/10.1016/j.tics.2012.02.002

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*, 441–480. http://doi.org/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. http://doi.org/10.1016/j.neuropsychologia.2012.08.001

Kail, R. (2000). Speed of Information Processing: Developmental Change and Links. *Journal of School Psychology*, *38*(1), 51–61.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding It Up: Helping Children Learn Mathematics*.

Lakoff, G., & Nunez, R. E. (2000). *Where Mathematics Comes From: How the embodied mind brings mathematics into being*. New York, NY: Basic Books.

Lamon, S. J. (2012). *Teaching fractions and ratios for understanding: Essential content knowledge and instructional strategies for teachers*. Routledge.

Lesh, R., & Yoon, C. (2004). Evolving Communities of Mind-In Which Development Involves Several Interacting and Simultaneously Developing Strands. *Mathematical Thinking and Learning*, *6*(2), 205–226.

Lewis, M. R., Matthews, P. G., & Hubbard, E. M. (2015). Neurocognitive Architecture and the Nonsymbolic Foundations of Fractions Understanding. In D. B. Berch, D. C. Geary, & K. Mann Knoepke (Eds.), *Development of mathematical cognition: Neural substrates and genetic influences* (pp. 141–164). Academic Press Inc. http://doi.org/10.1017/CBO9781107415324.004

Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology*, *50*(4), 315–353. http://doi.org/10.1016/j.cogpsych.2004.09.004

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.

Luna, B., Velanova, K., & Geier, C. F. (2008). Development of eye-movement control. *Brain and Cognition*, *68*(3), 293–308. http://doi.org/10.1016/j.bandc.2008.08.019

Mack, N. K. (1993). Learning Rational Numbers with Understanding: the case of informal knowledge. In T. P. Carpenter, E. Fennema, & T. A. Romberg (Eds.), *Rational Numbers: an integration of research* (pp. 85–105). Hillsdale, NJ: Lawrence Erlbaum Associates.

Mazzocco, M. M. M., & Devlin, K. T. (2008). Parts and "holes": gaps in rational number sense among children with vs. without mathematical learning disabilities. *Developmental Science*, *11*(5), 681–691. http://doi.org/10.1111/j.1467-7687.2008.00717.x

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. http://doi.org/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 (2006), *62*(8), 1598–1616. http://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. http://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. http://doi.org/10.1016/j.jecp.2010.04.008
Miller Singley, A. T., & Bunge, S. A. (n.d.). Strategy Adaptation in a Fraction Comparison Task: An eye-tracking study.
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. http://doi.org/10.1016/j.tine.2014.03.001
Moore, T., & Fallah, M. (2004). Microstimulation of the frontal eye field and its effects on covert spatial attention. *Journal of Neurophysiology*, *91*(1), 152–162. http://doi.org/10.1152/jn.00741.2002
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), 542–53. http://doi.org/10.1111/desc.12048
Moss, J., & Case, R. (1999). Developing Children' s Understanding of the Rational Numbers: A New Model and an Experimental Curriculum. *Journal for Research in Mathematics Education*, *30*(2), 122–147.
Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*(5109), 1519–1520. http://doi.org/10.1038/2151519a0
Muller, J. R., Philiastides, M. G., & Newsome, W. T. (2005). Microstimulation of the superior colliculus focuses attention without moving the eyes. *PNAS*, *102*(3), 524–529.
Munoz, D. P., Broughton, J. R., Goldring, J. E., & Armstrong, I. T. (1998). Age-related performance of human subjects on saccadic eye movement tasks. *Experimental Brain Research*, *121*(4), 391–400. http://doi.org/10.1007/s002210050473
Murata, A. (2013). Diversity and High Academic Expectations Without Tracking: Inclusively Responsive Instruction. *Journal of the Learning Sciences*, *22*(2), 312–335. http://doi.org/10.1080/10508406.2012.682188
National Council of Teachers of Mathematics. (2007). *Second handbook of research on mathematics teaching and learning.* Washington, DC, US.
National Mathematics Advisory Panel. (2008). *Foundations for Success: The Final Report of the National Mathematics Advisory Panel*. Washington, DC, US.
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. http://doi.org/10.1207/s15326985ep4001
Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annual Review of Neuroscience*, *32*, 185–208. http://doi.org/10.1146/annurev.neuro.051508.135550

Obersteiner, A., Moll, G., Beitlich, 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. http://doi.org/10.1007/s11858-015-0742-z

Piaget, J. (1970). Genetic Epistemology -Chapter 1. In *Genetic Epistemology*.

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 (2006)*, *61*(8), 1143–1150. http://doi.org/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. http://doi.org/10.1016/j.lindif.2010.05.001

Rattermann, M. J., & Gentner, D. (1998). The effect of language on similarity: The use of relational labels improves young children's performance in a mapping task. In K. J. Holyoak, D. Gentner, & B. Kokinov (Eds.), *Advances in analogy research: Integration of theory and data from the cognitive, computational, and neural sciences* (pp. 274–282). Sophia: New Bulgarian University.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. http://doi.org/10.1037/0033-2909.124.3.372

Rayner, K. (2009). *Eye movements and attention in reading, scene perception, and visual search. Quarterly journal of experimental psychology (2006)* (Vol. 62). http://doi.org/10.1080/17470210902816461

Reys, B., Dingman, S., Nevels, N., & Teuscher, D. (2007). High School Mathematics: State-Level Curriculum Standards and Graduation Requirements. *Journal of Mathematics Curriculum*.

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*, (4), 1–18. http://doi.org/10.1177/2372732216629795

Richland, L. E., Begolli, K. N., Simms, N., Frausel, R. R., & Lyons, E. A. (2016). Supporting Mathematical Discussions: the Roles of Comparison and Cognitive Load. *Educational Psychology Review*. http://doi.org/10.1007/s10648-016-9382-2

Ross, R. G., Radant, A. D., Young, D. A., & Hommer, D. W. (1994). Saccadic Eye Movements in Normal Children from 8 to 15 Years of Age: A Developmental Study of Visuospatial Attention. *Journal of Autism and Developmental Disorders*, *24*(4), 413–431.

Salman, M. S., Sharpe, J. A., Eizenman, M., Lillakas, L., Westall, C., To, T., … Steinbach, M. J. (2006). Saccades in children. *Vision Research*, *46*, 1432–1439. http://doi.org/10.1016/j.visres.2005.06.011

Saxe, G. B., De Kirby, K., Kang, B., Le, M., & Schneider, A. (2015). Studying cognition through time in a classroom community: The interplay between "everyday" and "scientific concepts." *Human Development*, *58*(1), 5–44. http://doi.org/10.1159/000371560
Saxe, G. B., DeKirby, K., Le, M., Sitabkhan, Y., & Kang, B. (2013). Understanding Learning across Lessons in Classroom Communities: A multi-leveled analytic approach. In A. Bikner-Ahsbahs, C. Knipping, & N. Presmeg (Eds.), *Advances in Mathematics Education*. ZDM.
Saxe, G. B., Diakow, R., & Gearhart, M. (2012). Towards curricular coherence in integers and fractions: a study of the efficacy of a lesson sequence that uses the number line as the principal representational context. *Zdm*, *45*(3), 343–364. http://doi.org/10.1007/s11858-012-0466-2
Saxe, G. B., Earnest, D., Sitabkhan, Y., Haldar, L. C., Lewis, K. E., & Zheng, Y. (2010). Supporting Generative Thinking About the Integer Number Line in Elementary Mathematics. *Cognition and Instruction*, *28*(4), 433–474. http://doi.org/10.1080/07370008.2010.511569
Saxe, G. B., Shaughnessy, M. M., Gearhart, M., & Haldar, L. C. (2013). Coordinating Numeric and Linear Units: Elementary Students' Strategies for Locating Whole Numbers on the Number Line. *Mathematical Thinking and Learning*, *15*(4), 235–258. http://doi.org/10.1080/10986065.2013.812510
Schneider, M., Grabner, R. H., Zurich, E., & Paetsch, J. (2009). Mental Number Line, Number Line Estimation, and Mathematical Achievement: Their Interrelations in Grades 5 and 6. *Journal of Educational Psychology*, *101*(2), 359–372. http://doi.org/10.1037/a0013840
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. http://doi.org/10.1016/j.cogdev.2008.07.002
Schneider, M., Rittle-Johnson, B., & Star, J. R. (2011). Relations among conceptual knowledge, procedural knowledge, and procedural flexibility in two samples differing in prior knowledge. *Developmental Psychology*, *47*(6), 1525–38. http://doi.org/10.1037/a0024997
Schneider, M., & Siegler, R. S. (2010). Representations of the Magnitudes of Fractions. *Journal of Experimental Psychology. Human Perception and Performance*, *36*(5), 1227–1238. http://doi.org/10.1037/a0018170
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. http://doi.org/10.1080/14640748608401609
Siegler, R. S. (2016). Magnitude knowledge: the common core of numerical development. *Developmental Science*, *19*(3), 341–61. http://doi.org/10.1111/desc.12395
Siegler, R. S., Thompson, C. a, & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology*, *62*(4), 273–96. http://doi.org/10.1016/j.cogpsych.2011.03.001
Simon, M. A. (1995). Reconstructing Mathematics Pedagogy from a Constructivist

Perspective. *Journal for Research in Mathematics Education*, *26*(2), 114–145.
Smith III, J. P. (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. http://doi.org/10.1111/j.1751-228X.2011.01109.x
Stafylidou, S., & Vosniadou, S. (2004). The development of students' understanding of the numerical value of fractions. *Learning and Instruction*, *14*(5 SPEC.ISS.), 503–518. http://doi.org/10.1016/j.learninstruc.2004.06.015
Starr, A., & Brannon, E. M. (2015). *Evolutionary and Developmental Continuities in Numerical Cognition*. *Evolutionary Origins and Early Development of Number Processing* (1st ed., Vol. 1). Elsevier Inc. http://doi.org/10.1016/B978-0-12-420133-0.00005-3
Sternberg, R. J. (1977). A componential theory of analogical reasoning. In A. W. Melton (Ed.), *Intelligence, information processing, and analogical reasoning* (pp. 134–172). Hillsdale, NJ: Lawrence Erlbaum Associates.
Stigler, J. W., Givvin, K. B., & Thompson, B. J. (2010). What Community College Developmental Mathematics Students Understand about Mathematics. *MathAMATYC Educ*, *1*(3), 4–16.
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. http://doi.org/10.3758/s13423-011-0081-1
Sztajn, P., Confrey, J., Wilson, P. H., & Edgington, C. (2012). Learning Trajectory Based Instruction: Toward a Theory of Teaching. *Educational Researcher*, *41*(5), 147–156. http://doi.org/10.3102/0013189X12442801
Thibaut, J., French, R., Missault, A., Gerard, Y., & Glady, Y. (2011). In the eyes of the beholder: What eye-tracking reveals about analogy-making strategies in children and adults. *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, 453–458.
Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language*, *53*(2), 204–224. http://doi.org/10.1016/j.jml.2005.02.010
U.S. Department of Education. Institute of Education Sciences. National Center for Education Statistics. (2015). *Digest of Education Statistics*. Washington, DC, US.
Vamvakoussi, X., & Vosniadou, S. (2004). Understanding the structure of the set of rational numbers: a conceptual change approach. *Learning and Instruction*, *14*, 453–467. http://doi.org/10.1016/j.learninstruc.2004.06.013
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. http://doi.org/10.1007/s00426-015-0736-z
Vendetti, M. S., Starr, A., Johnson, E. L., Modavi, K., & Bunge, S. A. (n.d.). Eyegaze Patterns Reveal Optimal Strategies during Analogical Reasoning. *In Revision*.
Vermunt, J. K., Langeheine, R., & Bockenholt, U. (1999). Discrete-Time Discrete-State Latent Markov Models with Time-Constant and Time-Varying Covariates.

Journal of Educational and Behavioral Statistics, *24*(2), 179–207. http://doi.org/10.3102/10769986024002179
Vermunt, J. K., & Magidson, J. (2002). Latent Class Cluster Analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89–106). Cambridge, UK: Cambridge University Press.
Vukovic, R. K., Fuchs, L. S., Geary, D. C., Jordan, N. C., Gersten, R., & Siegler, R. S. (2014). Sources of individual differences in children's understanding of fractions. *Child Development*, *85*(4). http://doi.org/10.1111/cdev.12218
Vygotsky, L. S. (1978). *Mind in Society: The development of higher psychological processes*. (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge, MA: Harvard University Press.
Wade, N. J. (2015). How Were Eye Movements Recorded Before Yarbus? *Perception*, *44*(8–9), 851–883. http://doi.org/10.1177/0301006615594947
Wendelken, C., & Bunge, S. A. (2010). Transitive inference: Distinct contributions of rostrolateral prefrontal cortex and the hippocampus. *Journal of Cognitive Neuroscience*, *22*(5), 837–847. http://doi.org/10.1162/jocn.2009.21226.Transitive
Wendelken, C., Chung, D., & Bunge, S. A. (2011). Rostrolateral prefrontal cortex: Domain-general or domain-sensitive? *Human Brain Mapping*, *0*. http://doi.org/10.1002/hbm.21336
Wendelken, C., O'Hare, E. D., Whitaker, K. J., Ferrer, E., & Bunge, S. a. (2011). Increased functional selectivity over development in rostrolateral prefrontal cortex. *The Journal of Neuroscience*, *31*(47), 17260–17268. http://doi.org/10.1523/JNEUROSCI.1193-10.2011
Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. *Journal of Research in Science Teaching*, *46*(6), 716–730. http://doi.org/10.1002/tea.20318
Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock-Johnson III Tests of Cognitive Abilities. Itasca, IL: Riverside Publishing.
Wright, V. J. (2011). *The Development of Multiplicative Thinking and Proportional Reasoning: Models of conceptual learning and transfer*. University of Waikato.

Appendix A

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	$\frac{6}{8}$ $\frac{5}{7}$	$\frac{5}{7}$ $\frac{6}{8}$	$\frac{2}{4}$ $\frac{3}{5}$	$\frac{3}{5}$ $\frac{2}{4}$	$\frac{4}{8}$ $\frac{3}{7}$	$\frac{3}{7}$ $\frac{4}{8}$	$\frac{2}{6}$ $\frac{3}{7}$	$\frac{3}{7}$ $\frac{2}{6}$

Table A1. List of Stimuli used in Fraction Comparison Task with Fifth Graders and Adults, as reported in Chapter 1

Appendix B

M+H+	$\frac{1}{3}$ $\frac{7}{12}$	$\frac{2}{5}$ $\frac{8}{15}$	$\frac{3}{7}$ $\frac{9}{14}$	$\frac{4}{7}$ $\frac{5}{14}$	$\frac{5}{9}$ $\frac{1}{3}$
	$\frac{5}{12}$ $\frac{2}{3}$	$\frac{5}{9}$ $\frac{7}{18}$	$\frac{7}{15}$ $\frac{3}{5}$	$\frac{7}{16}$ $\frac{5}{8}$	$\frac{11}{18}$ $\frac{4}{9}$
M+H-	$\frac{1}{5}$ $\frac{7}{15}$	$\frac{2}{3}$ $\frac{7}{9}$	$\frac{3}{8}$ $\frac{1}{4}$	$\frac{3}{10}$ $\frac{2}{5}$	$\frac{3}{4}$ $\frac{5}{8}$
	$\frac{4}{5}$ $\frac{7}{10}$	$\frac{4}{7}$ $\frac{9}{14}$	$\frac{7}{10}$ $\frac{3}{5}$	$\frac{7}{12}$ $\frac{5}{6}$	$\frac{7}{9}$ $\frac{11}{18}$
M-H+	$\frac{3}{8}$ $\frac{7}{12}$	$\frac{6}{13}$ $\frac{4}{7}$	$\frac{7}{16}$ $\frac{3}{5}$	$\frac{7}{13}$ $\frac{5}{11}$	$\frac{7}{13}$ $\frac{8}{19}$
	$\frac{7}{12}$ $\frac{9}{19}$	$\frac{7}{15}$ $\frac{11}{17}$	$\frac{8}{13}$ $\frac{5}{12}$	$\frac{11}{18}$ $\frac{3}{8}$	*
M-H-	$\frac{3}{11}$ $\frac{2}{9}$	$\frac{3}{14}$ $\frac{4}{15}$	$\frac{3}{7}$ $\frac{4}{9}$	$\frac{5}{17}$ $\frac{3}{7}$	$\frac{5}{9}$ $\frac{7}{13}$
	$\frac{7}{11}$ $\frac{5}{8}$	$\frac{8}{19}$ $\frac{5}{11}$	$\frac{9}{14}$ $\frac{7}{12}$	$\frac{5}{9}$ $\frac{9}{17}$	*

Table B1. List of Stimuli used in Fraction Comparison Task with Adults, as reported in Chapter 2

* Due to a coding error, two trials were misclassified into these categories. They were corrected during the analysis stage. Because all analyses were completed by trial, hierarchically within participant, this is just one reason for unbalanced data across the conditions. Other reasons were missed responses and lack of saccades of interest within a particular trial.

Note: This list represents the trials within one run of the experiment. For the other run, all pairs were switched left to right and re-ordered within the run.