Visual Flexibility in Arithmetic Expressions

Jingqi Yu jy45@ iu.edu
David Landy dhlandy@gmail.com
Robert L. Goldstone rgoldsto@indiana.edu
Psychological and Brain Sciences, Indiana university, Bloomington, IN 47405

Abstract

We investigated whether, and in what, ways people use visual structures to evaluate mathematical expressions. We also explored the relationship between strategy use and other common measures in mathematics education. Participants organized long sum/products when visual structure was available in algebraic expressions. Two experiments showed a similar pattern: One group of participants primarily calculated from left to right, or combined identical numbers together. A second group calculated adjacent pairs. A third group tended to group terms which either produced easy sums (e.g., 6+4), or participated in a global structure. These different strategies were associated with different levels of success on the task, and, in Experiment 2, with differential math anxiety and mathematical skill. Specifically, problem solvers with lower math anxiety and higher math ability tend to group by chunks and easy calculation. These results identify an important role for the perception of coherent structure and pattern identification in mathematical reasoning.

Keywords: numerical cognition; mathematical cognition; cognitive psychology; educational psychology

Introduction

Consider the arithmetic expression 1 + 2 + 3 + 4 + 1 + 2 + 3 + 4 + 1 + 2 + 3 + 4. What strategy would you use to calculate this sum (without a calculator)? Do you prefer to simply add the addends from left to right? Do you find this one-size-fits-all approach (assuming compliance with the order of precedence) less mathematically satisfying because of the visual structure in this expression? Does the repeated number value catch your attention and encourage you to do $1\times3+2\times3+3\times3+4\times3=3+6+9+12=30$? Or, alternatively, are you more sensitive to the three groups of 1+2+3+4 and decide to parse the expression according to these groups, eventually calculating $3\times(1+2+3+4)$. Because they can all lead to a correct answer, there is no right or wrong solution. Theoretically, the number of possible solutions for an arithmetic expression can be infinite (e.g., reorganizing a sequence in a self-preferred way), but some of them appear more sensible and common than others depending on the problem structure. For example, grouping same numbers is not applicable to expressions without repeated numbers. Nevertheless, many mathematical representations contain internal coherent structure, and that structure may be an important determiner of how mathematical reasoners do in fact choose to solve problems.

A Tale of Two Explanations

At the surface level, the strategy applied to evaluate an algebraic expression may seem secondary to the answer's accuracy. In school, students doing arithmetic worksheets are often graded based on their answers but not intermediate steps of their strategies. Oftentimes, teachers and students only go back to the solution when the calculation is wrong, hoping to find out where the mistake is. However, intuitively, many people, including us, do not think all solutions are created equal. Indeed, multiple "easy and quick" routes (aka shortcuts) have been proposed (e.g., Asimov, 1964). Rearranging an expression so that items can cancel each other out or make multiples of 10 feels clever, efficient, and insightful (Benjamin & Shermer, 2006). This is especially the case when the alternative strategy substantially reduces computation cost. These alternatives are usually cognitively driven. Despite different shortcuts, a common motivation is to reduce intrinsic cognitive load by grouping interacting elements into chunks (Sweller & Chandler, 1994; Paas et al., 2003).

Take the expression we started with as an example: 1+ 2 + 3 + 4 + 1 + 2 + 3 + 4 + 1 + 2 + 3 + 4, one solution path is to realize that (1+ 2 + 3 + 4) + (1 + 2 + 3 + 4) + (1 + 2 + 3 + 4) = 3 × (1+ 2 + 3 + 4) = 3 × 10 = 30. There are at least two ways to explain the rationale behind this — one is more cognitively inspired while the other one is more perceptually inspired. First, the solution exemplifies a specific mathematical trick – making multiples of 10. We imagine an individual looking at the problem with the hope of finding some numbers that add up to 10. Because 1+ 2 + 3 + 4 = 10, that individual chooses to segment the expression into three groups of 1 + 2 + 3 + 4. This solution would apply equally well to a problem such as 1 + 2 + 3 + 4 + 7 + 3, which contains two groups that sum to 10. However, the former problem also contains a coherent visual structure and layout. The three groups are identical, separate, and adjacent, all features of good grouping that underlie perceptual organization. In line with the idea that visuospatial organization guides algebraic (Kirshner, 1989; Landy & Goldstone, 2007) and arithmetic (Landy & Goldstone, 2010; Braithwaite et al, 2016; Landy & Goldstone, 2007b) reasoning in the specific case of order of operation, it is possible that this irrelevant visual structure also guides mathematical reasoning when calculating complex sums in which multiple pathways are equally valid and appropriate—but unequally clever and efficient.

These two explanations are by no means mutually exclusive, especially given that arithmetic strategies can be discovered without conscious awareness (Siegler & Stern, 1998). It is worth noting that we do not plan to identify which account is more accurate. Instead, we wish to explore whether and in what ways structure, when present, will be used by people to solve algebraic expressions.

How Perceptual Cues Shape Formal Reasoning

Making sense of symbols is one of the most powerful and fascinating cognitive traits that characterize humans. Most research into formal symbolic reasoning has focused on the symbolic and abstract aspects of formal symbol systems. The general conclusion is that symbolic reasoning depends on internal structural rules, but not external structural forms (e.g., Gentner, 2003). So visuospatial structure seems peripheral to symbolic reasoning. This is especially the case with mathematical reasoning, which is often considered a purely formal discipline. Still, evidence has shown that mathematical reasoning is strongly grounded in visual processing. There is a general relationship between physical and syntactic proximity in notational mathematics (Kirsher, 1989). For example, the exponent is always closer to the base than to other terms in an expression, echoing its high precedence in the order of operations. People are also sensitive to visuospatial features of abstract representations in addition to the mathematical contents (McNeil & Alibali, 2004). Recently, there is a growing amount of evidence that non-formal contexts influence higher-level reasoning in various fields, such as graph interpretation (e.g., Gattis & Holyoak, 1996), physics (Larkin & Simon, 1987), and even mathematics (Landy & Goldstone, 2007b). The visual system can be trained to grasp and appreciate affordances of different types of display formats and to extract task-relevant information from external representations (Goldstone, Landy, & Son, 2010; Kellman et al., 2008).

The Gestalt Principles of Perception in Mathematics

Research into the role of visuospatial features in mathematics has continuously gained momentum with the rise of perceptual training. An important extant question is whether and when perceptual organization facilitates mathematical reasoning with symbols, and when it acts as a crutch or a distraction (Goldstone, Landy, & Son, 2010; Kellman, Massey, & Son, 2010). Kirshner and Awtry (2004) argue that employing visual heuristics in mathematics may block learning of principles and formal rules, and so impair learning. Nogueira de Lima & Tall (2008) support this by observing that students with weaker understanding of algebraic transformations are more likely to invoke physical analogies over equations, such as 'moving' a term to the other side of an equation. Moreover, the alignment of surface features between equations and expressions has been identified as an important source of errors in novices and experts (Landy, Brookes, & Smout, 2014). Incorrect use of visual patterns often leads to misleading generalizations (Marquis, 1988; Nogueira de Lima & Tall, 2008). However, there is a long tradition in psychology, dating back to the Gestalt psychologists, of invoking perceptual organization as an important factor in reasoning and problem-solving (Ohlsson, 1984). Appropriate chunking of input information enables the discovery of higher-order invariances (Kellman et al, 2008). Recently, Braithwaite et al (2016) suggested that children importantly learn relations between visual and formal properties of mathematics. They note that sensitivity to spacing variations increased rather than decreased with age and suggested that perceptual learning provides one way to off-load mathematical rules onto perceptual systems so as to decrease load on executive processes. However, a direct link between mathematical achievement and sensitivity to perceptual organization has not been shown, nor have algorithms or strategies been identified that would take direct advantage of perceptual organization.

The Present Study

In the current work, we explore strategies involved in computing long arithmetic expressions. We manipulated operands' neighboring values to create different kinds of higher-order organizations (e.g., 4 + 4 + 6 vs. 4 + 6 + 4 with equal space between adjacent elements). Second, we went a step further by asking reasoners to compute arithmetic expressions (long sums/products) instead of judging the validity of equations. Research into metacognition suggests that adult problem solvers can monitor their progress and adjust their methods as they proceed if they realize that the problem is more complex than they had first thought (Schoenfeld, 1992). Thus, there are good reasons to think that the effect of perceptual groupings may differ between judging and computing arithmetic expressions because the latter is generally more cognitively demanding.

Experiment 1

Experiment 1 explored whether and how people use grouping pressure available in long sums/products to compute arithmetic expressions and whether they are associated with different levels of arithmetic success.

Participants We recruited 33 undergraduates at Indiana University, Bloomington in exchange for course credit.

Stimuli The stimuli were a set of 24 long sums and products questions composed of multiple operators and operands. Each problem was designed to afford distinguishable perceptual cues. Questions differed either in lower-level features (e.g., numerical values) or higher-level structure (e.g., repetition of a conglomeration of terms). Taking 1 + 2 + 3 + 4 + 1 + 2 + 3 + 4 + 1 + 2 + 3 + 4 as an example, grouping cues include: a) high-level relational exactness:1 + 2 + 3 + 4 + 1 + 2 + 3 + 4 + 1 + 2 + 3 + 4 (1+2+3+4 repeated three times), b) high-level featural exactness: 1 + 2 + 3 + 4 + 2 + 3 + 4 + 1 + 3 + 4 + 1 + 2 (sums of 1 to 4 presented in dissimilar orders), and c) lower-level identity: 1 + 1 + 1 + 2 + 2 + 2 + 3 + 3 + 3 + 4 + 4 + 4 (adjacent common value). Other times, we varied number values for the same arrangement (e.g., 17 + 23 + 17 + 23 + 17 + 23 + 17 + 23 and 99 + 101 + 99 + 101 + 99 + 101 + 99 + 101). This allowed us to explore to what extent perceivers would be influenced by the actual values.

Procedure Experiment 1 was a within-subjects design in which participants answered 24 questions. The dependent measures included subjects' strategy use and their accuracy. Subjects were asked to compute each of the questions in the problem set to the best of their ability, but more importantly, to demonstrate their arithmetic reasoning by either using mathematical notions (e.g., parentheses and lines) or descriptive language (e.g., *group all the 5s together*) for each problem with a pen or a pencil. No calculator was allowed. They were assured that their overall performance would not affect their credit. Subjects also reported their age and sex after they completed the problem set.

Strategy Encoding There is no standard coding system for evaluating complex arithmetic expressions, and so we developed an approach tailored for this project. The data from this experiment were used to create a set of eight non-exclusive properties of subject strategies. Each specific response could be coded as matching or failing to match each code. We began from a set of theoretically interesting categories, including left to right ordering, chunking, and easy calculation, and added other categories that were needed to capture what seemed to be major repeated themes.

1) (Group from) left to right: Strict left to right
2) (Group by) common numeral
3) (Group by) neighbors: Group spatially adjacent terms (may or may not be every term)
4) (Group by) pairs: Break the long sum/product into shorter adjacent pairs with *dissimilar* sums
5) (Group into) higher-order patterns: Group into conglomerations of (at least three) terms
6) (Group by) easy common sum/product: intermediate steps produce 5 or 10
7) (Group by) sign: Group every same operator together when there are mixed operators.
8) (Group into) sorted order: Reorder operands in an ascending or descending order

Note: One participant rewrote all questions vertically and then did calculations based on decimal place values.

We performed a binary coding for our subjects' responses. Each arithmetic question was translated into an eight-dimensional vector by applying above rules one at a time.

Responses were encoded based on what was available and what was shown in participants' answers. For example, for those who indicated 3+13+23+33+43–3–3–3–3–3 as 3 + 13 + 23 + 33 + 43 – (3 + 3 + 3 + 3 + 3), their responses were encoded both as *Neighbor, Numeral*, and *Sign.* Of course, the participant may only have noticed a subset of these properties; categories were based on conformance to the rules, not participants' (hidden) intentions.

Results & Discussion

Strategy Use Analysis We conducted a principal component analysis (PCA) on the distribution of strategies for each subject, followed by k-means clustering. We represented each subject by collapsing their performance on the 24 questions. Thus, each subject is an eight-dimensional vector, with each dimension representing an averaged value for each grouping pattern. This transformation led to 33 data points in an eight-dimensional space. The correlation matrix (Figure 1) exhibits a moderate to strong positive correlation between *Common sum/product* and *Higher-order patterns*, which are both negatively correlated with *Numeral*, *Sorted order*, and *Left to right* at varying levels. The strong correlation between *Numeral* and *Sorted order* is partially due to participants' tendency to group common numerical value, followed by reordering. *Pair* and *Neighbor* are also strongly (positively) correlated. This is not surprising because the definition of *Pair* implies grouping neighbors.

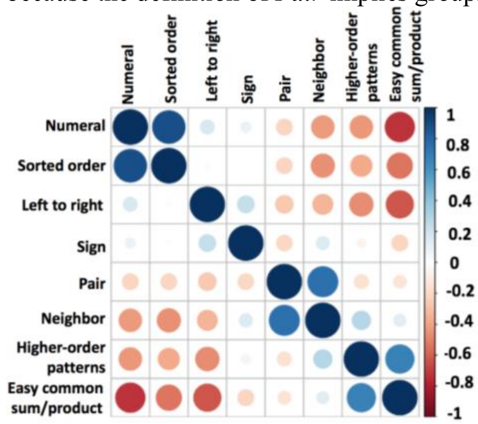

Figure 1. Correlation matrix for different grouping strategies

Three principal components were retained as they explained around 85% of variance altogether. Table 1 shows the factor loadings for grouping strategies in PCA. We identified three distinctive clusters by visual inspection. For simplicity, we named the clusters as follows: a) Surface Properties (N = 14): Go from left to right, sometimes clustering by numeric value or signs of operators, b) Near neighbors (N = 5): Group close entities, often into adjacent groups or pairs, and c) Higher-order patterns (N = 14): Group terms which were either led to multiples of 10, or which participated in a global structure (see Figure 2).

Table 1

Rotated factor loadings for grouping strategies

Items	Factor Loadings		
	PC1	PC2	PC3
Numeral	0.48	0.0048	-0.31
Left to Right	0.30	0.079	0.60
Sign	0.09	0.084	0.52
Pair	-0.18	0.67	-0.25
Neighbor	-0.34	0.53	-0.31
Higher-order patterns	-0.38	-0.33	-0.059
Easy common sum/product	-0.45	-0.38	-0.078
Sorted order	0.43	-0.074	-0.45

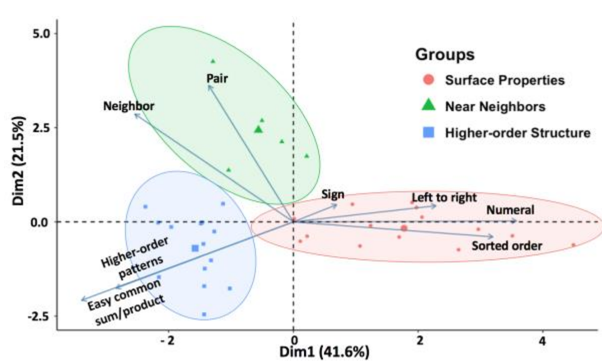

Figure 2. A PCA bi-plot of individuals and strategies with axes being the first two principal components. PC1 explains 41.6% of variance and PC2 explains 21.5% of variance

Accuracy by Cluster Subjects' response accuracy was also coded along with each strategy. Questions that were not given answers were coded as wrong. A failure to calculate an answer may be indicative of a poor choice of strategy, because the chosen strategy may involve too many calculations. Given three distinctive clusters of strategies, a natural question to ask is whether their accuracy in solving arithmetic expressions differs. We conducted an analysis of variance (ANOVA) to examine whether distinctive clusters led to different success rates. The analysis showed a significant association between the two, $F(2,30) = 7.314$, $p = .001$, $\eta_p{}^2 = .33$. Tukey's HSD post hoc test indicated that the average accuracy of *Higher-order Structure* ($M = .84$, $SD = .087$) was significantly higher than that of *Surface Properties* ($M = .66$, $SD = .18$), $p = .0039$. Similarly, the average of *Near Neighbors* ($M = .85$, $SD = .056$) was also significantly higher than that of *Surface Properties*., $p = .032$ (see Figure 3).

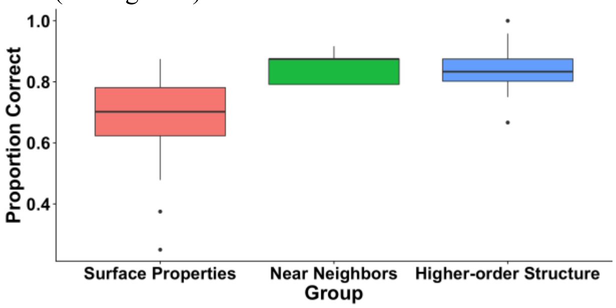

Figure 3. A boxplot of accuracy by cluster.

Overall, Experiment 1 showed that most people were inclined to use visual structures when they were available in long sums/products but that they used them differently. Different grouping strategies were observed and classified into three groups: *Surface Properties, Near neighbors*, and *Higher-order Structure.* Results also suggest that these distinctive clusters of strategy are associated with different levels of success, with people who use surface properties to evaluate arithmetic expressions having a lower rate of success than those in the alternative clusters.

Experiment 2

In light of the results of Experiment 1, we were particularly interested in the relationship between strategy use and common measures in mathematics education. Experiment 2 was designed to replicate and extend the findings of Experiment 1 by adding the Abbreviated Math Anxiety Scale (AMAS) (Hopko, Mahadevan, Bare, & Hunt, 2003) and the Berlin Numeracy Test (BNT) (Cokely et al, 2012). By including these two tests, it is possible to conclude whether the connection between strategy use and arithmetic success is only an artifact of our stimuli, or it in fact is associated with different levels of performance on some well-established measures in mathematics.

Participants We recruited 51 undergraduates at Indiana University, Bloomington in exchange for course credit.

Stimuli The arithmetic problems in Experiment 2 were similar to those in Experiment 1 except that there were 18 questions. The 9-item version of the AMAS was used to measure math anxiety. Participants rated how anxious they would feel during specified math-related events. Responses were on a Likert-type scale, ranging from 1 = *Low Anxiety* to 5 = *High Anxiety* (sample event: Listening to a lecture in mathematics class). The 4-item BNT was administered to assess statistical numeracy and risk literacy. A sample BNT question: Image we are throwing a five-sided die 50 times. On average, out of these 50 throws how many times would this five-sided die show an odd number (1, 3, or 5)? (30)

Procedure The procedure of Experiment 2 was identical to that of Experiment 1 except that participants were instructed to finish the BNT and AMAS after completing the arithmetic problems. In addition to age and gender, participants were also asked to report their math score on the SAT or ACT.

Results & Discussion

Strategy Use Analysis We retained 7 out of 8 strategies in Experiment 1, excluding *(group into) a sorted order* as it was not observed in Experiment 2. PCA and k-means clustering were applied to classify participants in terms of strategy use. The correlation matrix in Figure 4 reveals a pattern similar to the correlational finding in Experiment 1. Once again, we retained three principal components (Table 2) and observed three clusters: *Surface Properties*, *Near neighbors*, and *Higher-order Structure* (Figure 5).

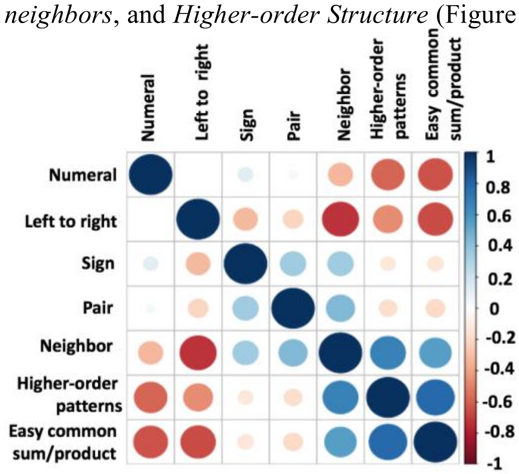

Figure 4. Correlation matrix for different grouping strategies

Table 2
Rotated factor loadings for grouping strategies

Items	Factor Loadings		
	PC1	PC2	PC3
Numeral	0.33	0.31	0.66
Left to Right	0.42	-0.28	-0.51
Sign	-0.068	0.56	0.088
Pair	-0.058	0.56	-0.53
Neighbor	-0.48	0.3	-0.083
Higher-order patterns	-0.49	-0.22	0.010
Easy common sum/product	-0.50	-0.24	0.14

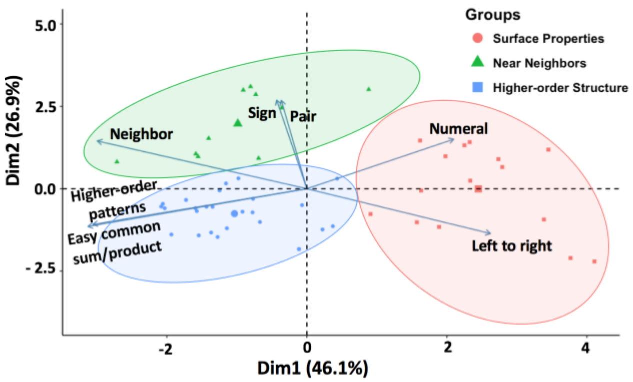

Figure 5. A PCA bi-plot of individuals and strategies with axes being the first two principal components. PC1 explains 46.1% of variance and PC2 explains 26.9% of variance.

Accuracy by Cluster An ANOVA indicated a significant effect of strategies on arithmetic success, $F(2,48) = 5.914$, $p = .0051$, $\eta_p{}^2 = .20$. Tukey's HSD post hoc test indicated that the average accuracy of participants in the *Higher-pattern Structure* cluster (M = .86, SD = .087) was significantly higher than that of *Near Neighbors* (M = .68, SD = .18), p = .0054. Nonetheless, the average of *Surface Properties (M =.76, SD = .15)* did not differ significantly from either group (see Figure 6).

Relationship to Measures in Mathematics We conducted an ANOVA for each measure.

Math Anxiety. The ANOVA showed that math anxiety level indeed associated with strategy use, $F(2, 48) = 9.068$, $p = .00046$, $\eta_p{}^2 = .27$. Tukey's HSD post hoc test indicated that participants in *Higher-order patterns* (M = 17, SD = 4.87) had a significantly lower level of anxiety than those in either *Surface Properties* (M = 24.07, SD = 5.79, p = .00054) or *Near Neighbors* (M = 22.1, SD = 6.05, p = .037). In other words, individuals with a lower level of math anxiety were more likely to group terms which either led to simple sums (especially multiples of 10) or participated in a global structure (see Figure 7).

Math Ability & Numeracy. We found reliable differences between different clusters on a few measures that are believed to be indicative of math ability and math anxiety. First, participants in distinctive clusters have significantly different math score on the SAT or ACT exam (transformed to the same percentage scale), $F(2,34) = 3.37$, $p = 0.046$, $\eta_p{}^2 = .17$ (14 participants were excluded for having not taken either test). Post hoc comparisons using the Tukey HSD test indicated that participants in *Higher-order Structure* (M = .82, SD = .13) had a significantly higher accuracy on high-stake standardized math tests than those in *Surface Properties* (M = .68, SD = .11), p = .04. A similar pattern was observed in their BNT scores, $F(2, 48) = 4.129$, $p = .02$, $\eta_p{}^2 = .15$ with those in *Higher-order Structure* (M = 2.4, SD = 1.42) scoring significantly higher on the BNT than those in *Surface Properties* (M = 1.2, SD = 1.15), p = .016. Figure 8 illustrates a striking trend that, in general, as BNT accuracy increases, the proportion of participants in *Higher-order Structure* increases while the proportion of participants in *Surface Properties* decreases. This is especially compelling given that the BNT and our task taps into different aspects of mathematical knowledge. These trends suggest an elegant connection between strategy and basic numeracy and mathematical literacy. By and large, participants who were better at using higher-order relations scored higher on the BNT.

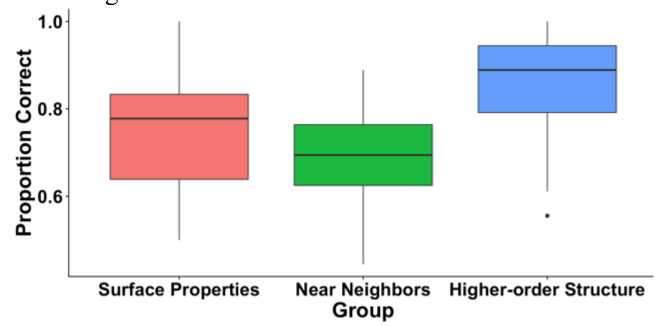

Figure 6. A boxplot of accuracy by cluster.

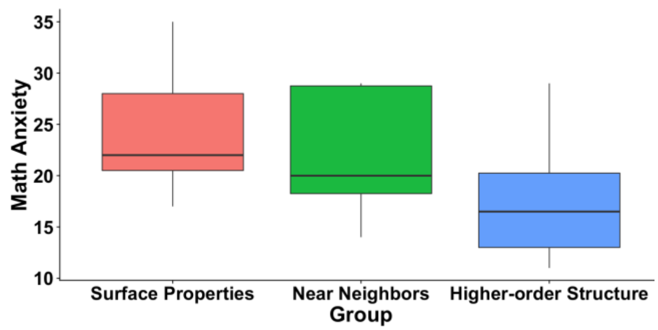

Figure 7. A boxplot of math anxiety by cluster (the higher the score the more mathematically anxious were participants).

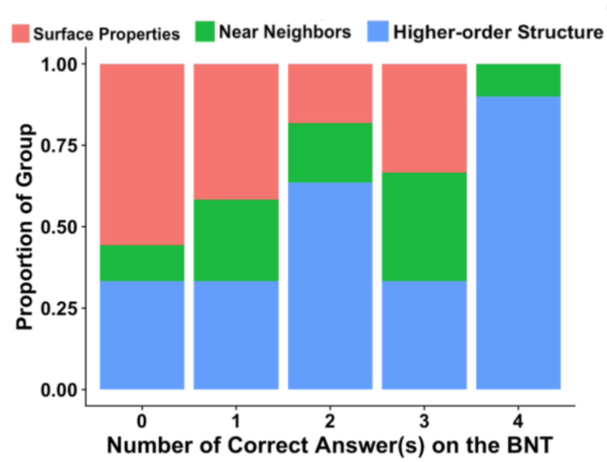

Figure 8. A stacked bar graph plots the proportion of subjects in each group at each level of BNT performance. The number of participants at each level is roughly equal: N = 9, 12, 11, 9, 10 at *BNT score* = 0, 1, 2, 3, 4, respectively.

Once again, the results of Experiment 2 supported the findings in Experiment 1 that people use visual structure to solve arithmetic expressions. Furthermore, Experiment 2 extended the findings of Experiment 1, revealing a clear relationship between grouping strategy and other common measures in mathematics education. Our results suggest that grouping by higher-order patterns is usually associated with lower math anxiety and higher BNT accuracy.

General Discussion & Conclusion

The present study, to the best of our knowledge, provides the first evidence that problem solvers use visual structural affordances to solve long sums/products. Across two experiments, we found that when visual structure was available, some participants relied on low-level properties (e.g., similarity between numerical value) while others tended to take advantage of higher-level patterns (e.g., exact pattern repetition). Some participants appeared to use fewer visual perceptual cues as they consistently shortened questions by pairing neighboring terms. We consider this difference in recognizing arithmetic patterns to reflect visual flexibility. People who tend to group math expressions into higher-order visual patterns are also likely to be able to solve problems by alternative strategies when necessary. Consequently, visual flexibility may directly impact procedural flexibility, which has been argued to be an important component of mathematical proficiency (Kilpatrick et al., 2001). However, the former, perceptual, type of flexibility has been less recognized and thus studied.

Patterning skill is considered by Steen (1988) to be at the core of mathematics. Many studies have explored relations between patterning skill and math ability, ranging from how young children identify and complete pattern extension tasks (e.g., Fyfe, McNeil, & Rittle-Johnson, 2015) to whether recurrent patterns can be observed through sets of analogies (Richland, Holyoak, & Stigler, 2004). Yet, little is known about patterning skill at the level of arithmetic expressions. While common core state standards for mathematics emphasize arithmetic pattern abstraction, the focus is on observations such as that the sum of two odd numbers is an even number. Our present study, in this sense, contributes to the literature on mathematics by showing that people also seek patterns at an arithmetic level. Moreover, individual differences in the kinds of arithmetic patterns spontaneously noticed is associated with measures of mathematics proficiency and anxiety. We consider visual flexibility and arithmetic pattern seeking to be an undervalued factor underlying skill in mathematics.

References

Asimov, I. (1964). *Quick and easy math.* Houghton Mifflin.

Benjamin, A., & Shermer, M. (2006). *Secrets of mental math: The mathemagician's guide to lightning calculation and amazing math tricks.* Three Rivers Press.

Braithwaite, D., Goldstone, R. L., van der Maas, H. L. J., & Landy, D. (2016). Informal mechanisms in mathematical cognitive development: the case of arithmetic. *Cognition, 149,* 40–55.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin Numeracy Test. *Judgment and Decision Making, 7,* 25–47.

Fyfe, E., McNeil, N. M., & Rittle-Johnson, B. (2015). Easy as ABCABC: Abstract language facilitates performance on a concrete patterning task. *Child Development, 86(3),* 927-935.

Gattis, M., & Holyoak, K. (1996). Mapping conceptual to spatial relations in visual reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1-9.

Gentner, D. (2003). Why we're so smart. In D. Gentner & S. GoldinMeadow (Eds.), *Language in mind: Advances in the study of language and thought*. Cambridge, MA: MIT Press.

Goldstone, R. L., Landy, D. H., & Son, J. Y. (2010). The education of perception. *Topics in Cognitive Science, 2(2),* 265–284.

Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. A. (2003). The Abbreviated Math Anxiety Scale (AMAS): Construction, validity, and reliability. *Assessment, 10,* 178-182.

Kellman, P. J., Massey, C. M., Roth, Z., Burke, T., Zucker, J., Saw, A., Aguero, K., & Wise, J. (2008). Perceptual learning and the technology of expertise: Studies in fraction learning and algebra. *Pragmatics & Cognition, 16(2),* 356–405.

Kellman, P. J., Massey, C. M., & Son, J. Y. (2010). Perceptual learning modules in mathematics: Enhancing students' pattern recognition, structure extraction, and fluency. *Topics in Cognitive Science, 2(2),* 285–305.

Kilpatrick, J., Swafford, J. O., & Findell, B. (2001). *Adding it up: Helping children learn mathematics.* Washington, DC: National Academies Press

Kirshner, D. (1989). The visual syntax of algebra. *JRME, 20(3),* 274–287.

Kirshner, D., & Awtry, T. (2004). Visual salience of algebraic Transformations. *JRME, 35(4),* 224.

Landy, D., & Goldstone, R. L. (2007b). How abstract is symbolic thought? *J Exp Psychol Learn Mem Cogn, 33*, 720–733.

Landy, D., & Goldstone, R. L. (2010). Proximity and precedence in arithmetic. *Q J Exp Psycho, 63,* 1953–1968.

Landy, D., Brookes, D., & Smout, R. (2014). Abstract numeric relations and the visual structure of algebra. *J Exp Psychol Learn Mem Cogn, 40,* 1404–1418.

Marquis, J. (1988). Common mistakes in algebra. In A. F. Coxford & A. P. Shulte (Eds.), *The ideas of algebra, K12: National Council of Teachers of Mathematics Yearbook.* Reston, VA: National Council of Teachers of Mathematics

McNeil, N. M., & Alibali, M. W. (2004). You'll see what you mean: Students encode equations based on their knowledge of arithmetic. *Cognitive Science, 28(3),* 451– 466.

Nogueira de Lima, R. and Tall, D. (2008). Procedural embodiment and magic in linear equations. *Educational Studies in Mathematics. 67 (1),* 3-18.

Ohlsson, S. (1984). Restructuring revisited: I. Summary and critique of the Gestalt theory of problem solving. *Scandinavian Journal of Psychology, 25,* 65–78.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38,* 1–4.

Richland, L. E., Holyoak, K. J., & Stigler, J. W. (2004). Analogy use in eighth-grade mathematics classrooms. *Cognition and Instruction, 22,* 37-60.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. Grouws (Ed.), *Handbook for research on mathematics teaching and learning* (pp. 334-370). New York: Macmillan.

Siegler, R. S., & Stern, E. (1998). Conscious and unconscious strategy discoveries: A microgenetic analysis. *Journal of Experimental Psychology-General, 127(4),* 377-397.

Steen B (1988) Body composition and aging. *Nutrition Reviews, 46, 45–51.*

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12,* 185–233.