How Space Guides Interpretation of a Novel Mathematical System

David Landy (dlandy@indiana.edu)
Department of Computer Science
Indiana University, Bloomington, IN 47408 USA

Robert L. Goldstone (rgoldsto@indiana.edu)
Department of Psychological and Brain Sciences
Indiana University, Bloomington, IN 47408 USA

Abstract

This paper investigates how people build interpretations of compound mathematical expressions in a novel formal system. In traditional arithmetic, interpretations are guided by an order of precedence convention (times and division precede addition and subtraction). This order is supported by a spatial convention that supports the order of precedence. In the experiment described here, participants learned computation tables of two simple novel operators, and then were asked to discover a precedence rule. The operators were presented with a physical spacing convention that either aligned with the precedence order, opposed it, or randomly opposed or aligned with the precedence order. Participants were more likely to reach a criterion of successful performance when the order of operations aligned with the precedence order, and did so more quickly than either other group. The results indicate that reasoners integrate salient perceptual cues with formal knowledge following familiar conventions, even on novel systems.

Keywords: Mathematical cognition, embodied cognition, formal reasoning, symbolic processing

Introduction

The ability to understand abstract formal structures is one of humanity's most distinctive and powerful cognitive traits. Arithmetical and algebraic notations, formal logic, and natural language syntax all contain underlying structure that at some level is entirely arbitrary and abstract. However, every actual notation has some particular physical presentation, and that notation always contains formally irrelevant physical relations. Often, especially when formal understanding is poor or partial, these relations may be more salient to a reasoner than the formally sanctioned abstract relations. The goal of this work is to explore whether and how people use irrelevant but salient visual information in exploring a novel formal system.

This issue has special importance for understanding mathematical reasoning and learning. Although arithmetic notation may be the best-known example of a purely formal symbol system, arithmetic itself contains a variety of non-formal conventions that relate visual aspects of expressions to their formal structure. One of the most striking of these is a correlation between physical proximity and the order of precedence in three common operators: addition, multiplication, and exponentiation. In the typeset expression

$$a + b \cdot c^2,$$

for instance, the *c* is closer to the exponent than to the *b*, and *b* is closer to the *c* than to the *a*. This physical relationship maps directly onto the corresponding order of precedence: *c* should first be squared, then the result multiplied by *b*, and *a* should be added to the result. The correspondence is far from perfect, especially when expressions are hand-written (Landy & Goldstone, in press A), but nevertheless, there is a general relationship between physical and syntactic proximity in mathematics, exacerbated by the frequent omission of the multiplication sign in algebra (of four prominent algebra textbooks (McGraw-Hill 1998; Cord Communications 2004; Holt, Rinehart and Winston, 2004; McDougal Littell, 2004), each used at least one times sign convention which was more closely spaced than plus signs, besides omission. None ever did the reverse).

People can learn and understand formal rule systems that lack the kind of perceptual-syntactic regularities algebra contains. The question addressed here is whether and in what ways salient perceptual regularities, when present, are used by people learning novel formal structures.

Background Kirshner (1989) explored the correspondence between spatial proximity and arithmetic syntax by creating a novel but natural set of symbols for the basic operations: M for multiplication, A for addition, E for exponentiation, and so on. Problems expressed in this language were presented with the natural spacing relationships to one group of high-school students, uniformly spaced to another. The participants who solved problems presented in the spaced language made many fewer errors compared to participants solving unspaced problems.

Kirshner's study demonstrates that arithmetic learners represent spacing regularities inherent in standard symbolic notation, and that they rely on the presence of those cues when developing interpretations of symbolic mathematical expressions. This is compatible with other studies that have shown that rule-based behavior uses irrelevant features of exemplars (McNeil & Alibali, 2004). One limitation of this work, however, is its use of standard arithmetic operations in the novel language. Because the stimuli are standard mathematical operations, it is difficult to determine the generality of the visual processes that govern order of operations behaviors. Participants have extensive experience with spacing in standard arithmetic symbology; this experience may drive their behavior when learning novel symbols for familiar operations without implying any general connection between syntax and spacing.

Studies of rule-based categorization have indicated that familiarity of contextual features can influence judgments, even when those features are known to be irrelevant (Allen & Brooks, 1991; Palmeri, 1997). Since Kirshner's stimuli map to the familiar operations, it is impossible to separate the effects of spacing familiarity from the role that spacing may play in guiding abstract interpretation generally. The current experiment expands on previous research by exploring the behavior of learners trying to understand a novel (generally mathematical) formal structure. Because the system is novel, familiarity and structural alignment effects can be cleanly separated.

In the domain of artificial grammar learning, Pothos (2005) explored the role of irrelevant variation on learning by manipulating the case of letters in stimulus sentences. Despite instructions to ignore the case of the letters, accuracy was lower when case was manipulated than in a single-case control. This study demonstrates that irrelevant variation can impact rule learning, but that variation makes the task uniformly more complex. In the study reported here, on the other hand (as in the case of algebraic equations), irrelevant variation is expected to simplify the task, by providing addition cues to structure.

Several interesting questions that can be asked about such a study include the following: Is the relationship between spacing and syntax applicable only to operations in which it has been learned, or will such a convention transfer to novel systems? If the latter, are these broader practices contingent and historical, or are they driven by underlying cognitive pressures? Will any kind of salient perceptual cue help? Finally, assuming that visual cues can improve performance, will those perceptual cues act as crutches, limiting or harming performance when spatial alignments are absent? Answers to these questions would inform cognitive theories of symbol learning, as well as having implications for mathematics education research and mathematical cognition. The experiment presented here provides an exploration of these issues, by asking participants to learn a novel pair of mathematical operations, and discover an order of operations rule governing them.

This novel system is presented to participants in one of three conditions: *aligned*, *inverted*, and *random*. The aligned condition is like the standard mathematical operators in that high-precedence operations are closely spaced. The inverted condition also provides a visual cue to precedence, but in this case the higher order operators are placed further apart. In the random condition, operators are randomly spaced narrowly or widely on each trial. In this last case, spacing variations—though present—are entirely uninformative.

There are two likely ways that reasoners might integrate spatial information in making perceptual judgments. If the primary advantage of spatial-syntactic regularities is the salient visual cue to structure, then randomly spaced structures should be harder than either aligned or inverted; since the former provides no visual information, while both aligned and inverted trials present salient 100%-valid cues to structure. Alternately, if the broader arithmetic practice of aligning short distances in additions, multiplications, and exponents with their order of precedence is a generalizable convention, then aligned trials should be quite easy, compared to both inverted and random trials.

The final stage of the experiment tests the robustness of the knowledge acquired, by removing spacing regularities. This phase is intended to evaluate whether spatial information that leads to correct judgments impedes subsequent understanding, as has been proposed, e.g., by Kirshner & Awtry (2004). Goldstone & Son (2005) argue that concrete trials presented in early training can support learning abstract concepts. The unsupported test phase presents one version of "concreteness fading"; valid visual cues to structure are removed. If these visual cues are used as crutches to replace syntactic knowledge, then performance should be roughly equal across all conditions, or even worse on conditions that show a benefit in the double-operator phase. If on the other hand these visual cues help guide syntactic understanding, then having experience with beneficial visual cues should lead to overall greater success even when support is removed.

Experiment

Method

68 Indiana University undergraduate students participated in this study for partial course credit. Of these, 5 were eliminated because they failed to reach criterion in the initial single-operator training stage, leaving 63 participants whose data were analyzed.

Participants learned two novel operations in isolation, and then had to discover a rule for how to combine them. The participants were instructed that the rule would be a simple order of precedence between the operators—one operator was to be bound before the other.

Single-operator training The experiment began with a single-operator training stage. In this phase, two novel operations, designated by the signs ֍and ۵, were defined over the symbols 0, 1, and 2 (see Tables 1 and 2). These operators were intended to look and feel mathematical, without reminding participants of any particular known operation, and to be balanced across response categories, and to be largely non-associative. The full operator tables for both operations were presented to participants before beginning the experiment, and after each section.

Table 1: The definition of the ֍ operator.

֍	0	1	2
0	2	2	1
1	2	1	0
2	1	0	0

Table 2: The definition of the Δ operator.

Δ	0	1	2
0	0	1	2
1	1	1	0
2	2	0	2

Single-operator training consisted of three sections: each section consisted of forced-choice trials, in which a single entry of a operator table was presented on a computer screen (for example, "1 ⊛ 1"). The stimulus remained until the participant responded by pressing a key corresponding to 0, 1, or 2. In section one, all trials involved one of the two operators; in section two, only the other operator appeared. The third section of single-operator learning contained trials with each of the two operators (though never both together in a single trial). Each section continued until a criterion of ten consecutive correct trials was reached, or until 300 trials were presented without reaching criterion. Participants failing to reach criterion in single-operator training have been removed from analysis. At least 30 trials were always presented in each section before the participant was allowed to proceed, in order to guarantee that participants had some time to familiarize themselves with the operations.

Double-operator stage The second part of the experiment presented compound problems in which both operations appeared in each expression. For instance, a participant might see the stimulus "1 Δ 2⊛1." The participants were instructed both at the beginning of the experiment, and immediately before double-operator trials began, that they would have to infer the rule for combining operators, but that one operator would be higher precedence than the other. In the example given, if ⊛ precedes Δ, then 1 Δ 2 ⊛ 1 reduces to 1 Δ 0, which reduces to 1, so the answer is 1. If, on the other hand, Δ precedes ⊛, then 1 Δ 2 ⊛ 1 reduces to 0 ⊛ 1, which reduces to 2.

Participants were presented with random problems, and made forced-choice responses, as in single-operator training. Which operation had higher precedence was counterbalanced across conditions. Once again, participants were tested until they reached a criterion of 10 in a row correct, or until they attempted 300 trials. These operators are non-associative, but imperfectly. This makes the task much more difficult, because participants received partial reinforcements for incorrect rules. It also makes the trials-to-criterion measure slightly less precise than might be hoped, since a participant might answer 10 problems in a row correctly despite using the reverse of the correct rule.

Throughout both single- and double-operator trials, operators were differentially spaced. Participants in the aligned condition always saw the higher-order operator spaced more narrowly than the secondary operator. For instance, if ⊛ precedes Δ, then a participant in the aligned condition would see problems like "1⊛2" and "0 Δ 0" in the single-operator training, and "1⊛0 Δ 0" in the double-operator phase. In the inverted condition, these regularities were reversed: the higher-order operator was always more widely spaced. In the random condition, spacing was randomized for each trial, with the constraint that on double-operator trials the operators were never presented with identical (both wide or both narrow) spacing.

Unsupported stage In the final stage of the experiment, participants solved problems which were formally identical to those of the double-operator phase, but spatial consistency was removed. In this phase, every trial was spaced randomly with the higher order operator either widely spaced, narrowly spaced, or with both operators spaced identically. Again, participants were tested until they reached a criterion of ten adjacent correct trials.

Results

Reaching criterion on this task proved extremely difficult. Of the 63 participants who successfully learned the meanings of single operators, only 34 (53%) mastered both the double-operator and unsupported stages. Participants in the different conditions fared differently (see Table 3): specifically, a higher proportion of participants reached criterion in the aligned than the inverted trials (72% vs. 34%, χ^2=4.58, $p<0.05$). Success reaching criterion in random trials did not differ from either aligned or inverted conditions.

Table 3: Number of participants reaching criterion on the double-op and unsupported experimental sections

	Condition		
Performance	Aligned	Inverted	Random
Reached Criterion	16	8	10
Did not Reach Criterion	8	17	8
% Successful	72%	34%	56%

Double-operator stage

Participants in the three conditions who reached criterion on all trials also differed in how many trials it took to reach that criterion. The mean number of trials taken to reach the end of each stage for each condition are presented in Table 4. While the single-operator training stage took participants in each condition roughly similar numbers of trials, the double-operator stage was mastered substantially faster by participants in the aligned condition than in either the

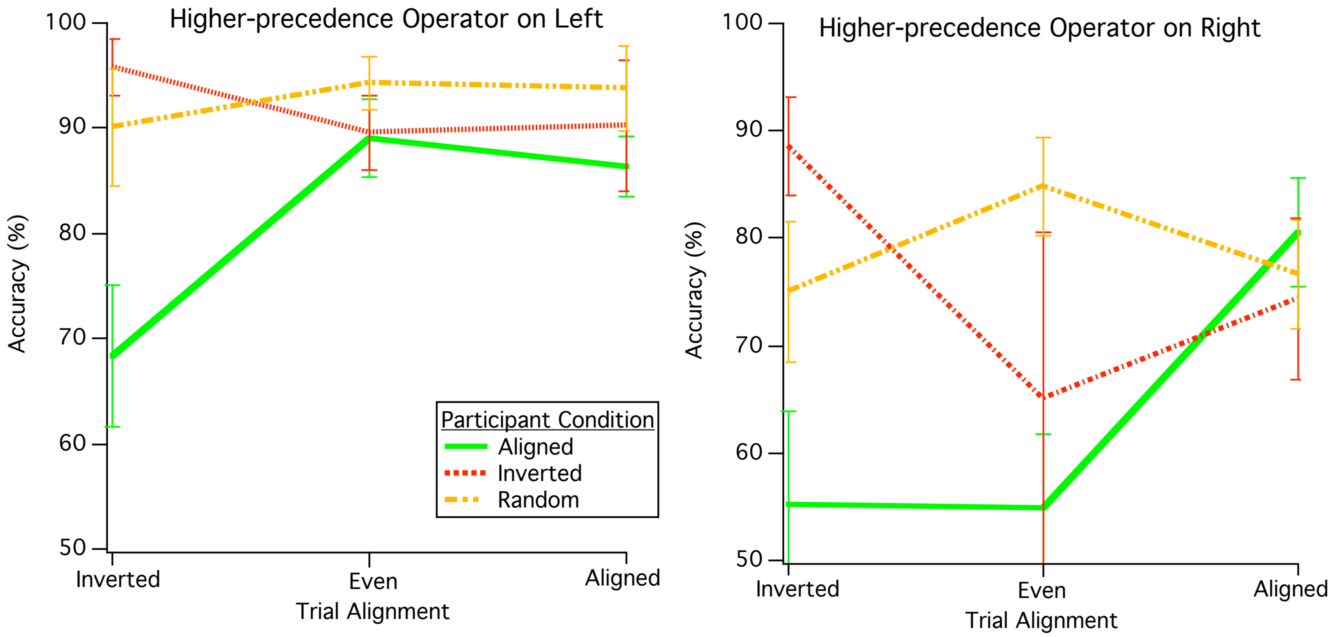

Figure 1: Mean accuracy in the unsupported stage vs. trial type for each of the three conditions, when the higher-order operator was on the left (left) and on the right (right). Generally, accuracy is higher when the high-order operator appeared on the left, and higher when the trial type matched the training condition.

random (t(24)=2.98, p<0.01) or the inverted conditions (t(22)=2.38, p<0.05).

Evidence for an alignment bias can also be seen in the behavior of participants in the random condition on individual trials. Since half of all trials in this condition have their spacing aligned with syntax, and half are inverted, differences in the accuracy on these trials provides an alternate measure of the alignment assumption. Table 5 presents the mean accuracy for the random condition, divided into trials in which the higher-order operator appeared on the left and on the right. As is indicated Table 5, and was verified by a 2-way ANOVA analysis, participants solved substantially more aligned than inverted trials, (81% vs. 61%, F(1,9)=7.3, p<0.02). The effect of operator position had a marginally significant effect on performance (74% vs. 69.6% accuracy, F(1,9)=5.2, p~.056).

Unsupported stage

The removal of spatial regularities hurt most those who gained the most from them. Participants in the aligned condition took substantially longer than in the inverted or random conditions to reach criterion in the unsupported stage (aligned vs. inverted t(22)=2.63, p<0.05; aligned vs. random t(20)= 2.62, p<0.05). The inconsistent and random conditions did not differ significantly.

Table 4: Mean number of trials to criterion (trials), with standard errors

	Condition		
Experimental Stage	Aligned	Inverted	Random
Single-op Training	139±17	131±13	136±24
Double-op stage	26±4	52±10	62±11
Unsupported stage	50±10	21±4	22±4

Table 5: Mean accuracy in the random condition on the double-operator stage, divided by position and spacing of high-order operator.

	Spacing	
Position	Aligned	Inverted
Left	84.3±4.4	64.4±7.6
Right	78.8±5.8	57.1±7.0

Since all three conditions contained trials that were aligned, inverted, and evenly spaced in the unsupported stage, an analysis of accurate trials by type is possible in all three conditions. The results are displayed in Figure 1. We performed a 3-way ANOVA using accuracy as the dependent measure, condition as a between-participants factor, and spacing alignment and the ordinal position of the higher-order operator as within-participants factors. This analysis revealed that mean accuracy was lower in the inverted condition than in the other two (73% vs. 84 and 86%, F(2, 31) =5.7, p<0.01). Also, accuracy was substantially higher when the left-most operator was higher-precedence (85% against 68.5%, F(1, 33) = 48.6, p<0.001). Trial alignment also had a main effect on accuracy; aligned stimuli were solved most successfully, and inverted trials least (70.4%, 76.2%, 83.6%, F(2, 65) = 8.17, p<0.001).

Despite the overall benefit of alignment, evidence can also be found for at least some types of familiarity. Participants in both the aligned and inverted conditions, were more accurate on trials which were familiar (i.e., that followed the spacing convention of the training phase) than those which were not. According to individual within-participants t-tests, participants in the aligned condition were substantially more accurate on aligned stimuli (t(15)=2.7, p<0.05). Despite the general advantage of aligned stimuli, participants in the inverted condition were marginally more accurate on familiar (inverted) stimuli than novel stimuli (t(7)=2.28, p=0.057) Participants in the random condition showed no benefit for familiar stimuli ($t(9)=0.8$, $p>0.4$).

Discussion

Spacing regularities informed syntactic judgments in this experiment, but only when that spacing aligns with common mathematical practice by placing higher-order operands together. Non-formal correspondences—even though they were highly salient and 100% valid—did not help participants determine order of precedence over no information when that correspondence violated the usual convention that closer spacing accompanies higher precedence. In contrast, participants learned the correct order much more quickly when it was aligned with a spatial correspondence. By and large, this evidence supports the hypothesis that the availability of a visual structure as a cue is mediated by the relationship between the structure and the formal structure it is aligned with. However, during the unsupported phase consistency was most helpful to participants in the consistent condition. This interaction suggests that participants did not just depend blindly on alignment, but accommodated to the local regularities to some degree.

Participants in the aligned condition took more trials than those in the random or inverted to master the same language once visual support was removed, indicating that to some degree these participants are using visual support as a crutch. However, in this case the crutch clearly supported eventual independent learning, since substantially more participants eventually learned the rule in the consistent condition.

The results demonstrate that the alignment between syntactic structure and spacing orthography is not restricted to the familiar mathematical operations, but is a general part of how people engage with mathematical structures structures. Although stimuli with familiar spacings may be easier to process than unfamiliar ones, this effect cannot explain the general alignment advantage shown here.

Possible sources of the alignment advantage

Unfortunately, one of the most interesting aspects of the alignment advantage is not addressed by this experiment: where does it come from? There are three plausible answers to this question. First, the alignment advantage seen in this experiment may well be a result of far transfer from the usual statistics of the familiar mathematical domain. In turn, spacing conventions in mathematics may be learned in each individual, and the biases seen in other studies (Kirshner, 1989; Landy & Goldstone, in press B) may result simply from that learning. This statistical account is in a certain sense unsatisfying: it might just as well have been the other way, that wider gaps would imply higher-order operations, had the orthographic choices of the original symbolic mathematicians been different. Being unsatisfying of course does not make this account less plausible. Another possibility is that the alignment of space and syntax tells us something deep about the mechanisms of learning formal syntax. It might be that formal syntax is, in some way, derived from the mechanisms that perform perceptual groups, in the same way that temporal language and judgments seem to be metaphorically derived from spatial judgments (Boroditsky, 2000). In this account, the observed alignment advantage is a trace of methods through which learners came to understand syntax. In skilled reasoners, however, syntax is processed using formal rule systems.

A final hypothesis is that the mechanisms used to process syntax are not, entirely, the symbolic mechanisms used to learn truly unsupported formal symbol systems, but are rooted in perceptual-motor systems that use visual cues to engage with mathematical texts as scenes. For instance, it may be that skilled mathematical reasoners pick out and attend first to closely spaced items, rather than reading equations from left to right. In helpfully spaced equations, such a process would obviate the need to represent a parse derived from operator order; in an unhelpfully constructed system, such as our inverted condition, this mechanism would backfire. What is interesting about this explanation is that it is not rooted in a statistical observation or belief ("Close items ought to bind more tightly"), but in a plausible computational *practice*. The apparent belief falls out of the way people engage with formal texts.

Sfard & Linchevski (1994) discuss the historical explosion of mathematics that accompanied the creation of modern symbolic algebra in the 15th century (previously, algebraic forms were written out in sentences as algorithms (Cajori, 1927)). Sfard & Linchevski suggest that one of the advantages of formal notations is that they allow users to treat as objects what seem to be processes. For instance, they suggest that it is easier to engage with "a + b*c" than "b multiplied by c, with the result added to a" as a *thing*. This perspective accords naturally with the process-oriented account of the previous paragraph. In this account, natural visual parsing cues are used to divide expressions up into their (visual) parts; these parts are treated as things, and similarly subdivided. As long as the visual segments align with the syntactic ones, object segmentation systems will automatically generate correct formal parsings (see also Endress, Scholl & Mehler, 2005).

The experiments presented here do not resolve the source of the alignment advantage. Dissociating the effects of experience with aligned notations, derivation of syntactic structure from spacing, and process-driven advantages for alignment will require future research. Regardless of the source of the advantage, the presence of a general relationship between syntactic structure and spacing has some general implications for both cognitive science and mathematical education.

Possible implications of the alignment advantage

People integrate spatial information implicit in the visual presentation of formal notations. Furthermore, this integration supports correct formal practice, when orthographic practices align with syntactic hierarchies. Thus, we suggest that visual and non-formal processes are substantially responsible for successful behavior in formal reasoning where such information is available. We do not think this is a radical, or even a very surprising conclusion, but it stands in stark contrast to extant claims that physical relationships other than concatenation are irrelevant to expression interpretation (Chandrasekaran, 2006; Stenning, 2002), or that such physical relationships are damaging and should be removed or attention to them discouraged (Iverson, 1980; Kirshner & Awtry, 2004). This suggestion is also incompatible with the standard practice of formal arithmetic modeling, which tends to ignore aspects of vision beyond basic symbol detection (e.g., Anderson, 2005), and of studies in mathematical cognition, which typically do not even report the spacing of presented stimuli (Koedinger & Nathan, 2004; Butterworth et al, 2001). The main implication of this work is that small variations in how formal terms are laid out on a page have large effects on how those terms are used by reasoners.

A second implication is that designers of novel languages would be well-served by a consideration of the general alignment of non-formal and formal regularities implicit in their systems. Kirshner & Awtry (2004) recommend that, because using visual similarity as a guide to formal arithmetic is sometimes misleading, students should be discouraged from using them at all. We feel, though, that the fault lies in our systems, not in ourselves. Systems that align these properties are likely to be substantially easier to learn and use than systems which do not.

Finally, this research has implications for cognitive scientists interested in abstract pattern learning. The explicit goal of many such researchers is to explore a fundamental abstract capacity to learn rule-governed systems (e.g., Marcus, 2001). Such research is valid and interesting, but it may be that the role of such systems for learning abstract patterns is not as large as has been assumed. Although genuinely abstract formal systems without perceptual cues may possibly be designed, the processes people use to successfully master real formal systems extend well beyond pure symbolic reasoning.

Acknowledgments

This research was funded by Department of Education, Institute of Education Sciences grant R305H050116, and National Science Foundation ROLE grant 0527920.

References

Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General*, *120*, 3-19.

Anderson, J.R. (2005). Human symbol manipulation within an Integrated Cognitive Architecture. *Cognitive Science 29*, 313-341.

Boroditsky, L.(2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition*, *75* (1), 1-28.

Butterworth, B., Zorzi, M., Girelli, L, Jonckheere, A.R. (2001). Storage and retrieval of addition facts: The role of number comparison. *The Quartery Journal of Experimental Pychology 54A*(4), 1005-1029.

Cajori, F. (1927). *A history of mathematical notations*. Lasalle, Illinois: The Open Court Publishing Co.

Chandrasekaran, B. (2006). What makes a bunch of marks a diagrammatic representation, and another bunch a sentential representation? *Proceedings of the AAAI Spring Symposium 2005, Reasoning with Mental and External Diagrams: Computational Modeling and Spatial Assistance.* Stanford University, California.

Cord Communications (2004). *Algebra I.* Waco, TX: Author.

Endress, A.D., Scholl, B.J., Mehler, J. (2005). The Role of Salience in the Extraction of Algebraic Rules. *Journal of Experimental Psychology: General 134*(3), 406-419.

Goldstone, R. L., & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *Journal of the Learning Sciences 14*, 69-110.

Holt, Rinehart & Winston (2004). *Algebra I.* Austin, TX: Author.

Iverson, K. E. (1980). Notation as a tool of thought. *Communications of the ACM, 23*(8), 444-465.

Kirshner, D. (1989). The visual syntax of algebra. *Journal for Research in Mathematics Education*, *20*(3), 274-287.

Kirshner, D., & Awtry, T. (2004). Visual salience of algebraic transformations. *Journal for Research in Mathematics Education, 35*(4), 224-257.

Koedinger, K. R. & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *Journal of the Learning Sciences*, *13*(2), 129-164.

Landy, D. & Goldstone, R. L. (in press A). Formal notations are diagrams: evidence from a production task. *Memory & Cognition*.

Landy, D. & Goldstone, R.L. (in press B). The alignment of order and space in arithmetic computation. *Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society*, 382-387. Hillsdale, NJ: Lawrence Erlbaum Associates.

Marcus, G. F. (2001). *The algebraic mind: Integrating connectionism and cognitive science*. Cambridge, MA: MIT Press.

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*, 451-466.

McDougal Littell (2004). *Algebra I.* Evanston, ILL: Author.

McGraw-Hill (1998). *Algebra 1*. Westerville, OH: Author.

Palmeri, T.J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning Memory & Cognition*. *23*, 324-354.

Pothos, E. M. (2005). Expectations about stimulus structure in implicit learning. *Memory & Cognition*, *33*, 171-181.

Sfard, A. & Linchevski, L. (1994). The gains and the pitfalls of reification: The case of algebra. *Educational Studies in Mathematics, 26*, 191-228.

Stenning, K. (2002). *Seeing Reason: Image and Language in Learning to Think*. Oxford University Press, Oxford.