Journal of Experimental Psychology:
Learning, Memory, and Cognition
2007, Vol. 33, No. 4, 720–733


0278-7393/07/$12.00 DOI: 10.1037/0278-7393.33.4.720

# How Abstract Is Symbolic Thought?

David Landy and Robert L. Goldstone
Indiana University

In 4 experiments, the authors explored the role of visual layout in rule-based syntactic judgments. Participants judged the validity of a set of algebraic equations that tested their ability to apply the order of operations. In each experiment, a nonmathematical grouping pressure was manipulated to support or interfere with the mathematical convention. Despite the formal irrelevance of these grouping manipulations, accuracy in all experiments was highest when the nonmathematical pressure supported the mathematical grouping. The increase was significantly greater when the correct judgment depended on the order of operator precedence. The result that visual perception impacts rule application in mathematics has broad implications for relational reasoning in general. The authors conclude that formally symbolic reasoning is more visual than is usually proposed.

*Keywords:* symbolic processing, mathematics, embodied cognition, relational reasoning, perceptual grouping

How thinking creatures manage to think using symbols is one of the central mysteries facing the cognitive sciences. Most research into formal symbolic reasoning emphasizes the abstract and arbitrary quality of formal symbol systems (Fodor, 1975; Gentner, 2003; Harnad, 1990; Haugeland, 1985; Jackendoff, 1983; Markman & Dietrich, 2000a, 2000b; Sloman, 1996). Symbolic reasoning is proposed to depend on internal structural rules, which do not relate to explicit external forms (e.g., Harnad, 1990; Markman & Dietrich, 2000a, 2000b; this perspective is also taken specifically with regard to notational mathematics in Stylianou, 2002; Zazkis, Dubinsky, & Dautermann, 1996). Mathematical and especially algebraic reasoning is often taken to be the paradigmatic case of pure symbolic reasoning, and to rely for its successful execution on the use of internally available formal operations (Inhelder & Piaget, 1958).

Although algebra probably is the best example we have of pure symbolic manipulation, it may not be a very good example. Although notational mathematics is typically treated as a particularly abstract symbol system, it is nevertheless the case that these notations are visually distinctive forms that occur in particular spatial arrangements and physical contexts. These special symbols are seen far more often in certain physical patterns than in others and are usually set off from the rest of a page or text. There is some evidence that people are sensitive not just to the contents but also to the particular perceptual forms of the representations of abstract entities when performing numerical calculations (Campbell, 1994; Zhang & Wang, 2005) and when interpreting algebraic expressions (McNeil & Alibali, 2004, 2005).

Previous research has also demonstrated the importance of nonformal context specifically on formal reasoning. In the domain of mathematics, Bassok (1997, 2001; Bassok, Chase, & Martin 1998; Wisniewski & Bassok, 1999) has demonstrated that computation problems are more readily solved, more easily constructed, and less likely to generate surprise when operations are semantically aligned with the typical structure of their contents. In deductive reasoning, a wide variety of research (for a review, see Johnson-Laird, 2001) has demonstrated the importance of nonformal figural context, such as the order of premises in a syllogism, on subsequent reasoning (Morley, Evans, & Handley, 2004). Tukey (1972) and Tufte (1983) have discussed ways that the visual details in the presentation of graphical data affect how people interpret those data. More generally, it has been proposed that many abstract concepts are processed through concrete, spatial metaphors (Boroditsky, 2000; Lakoff & Nũnez, 2000). The current research extends these demonstrations of the context specificity of formal reasoning by attending specifically to the role of the perceptual properties of notations in guiding reasoning. The bulk of research on physical grounding of mathematics has contrasted a grounded understanding of mathematics with ungrounded formal understandings. Our argument is that physical grounding is important even when reasoners are processing symbolic notations. Thus, rather than contrasting symbolic and grounded accounts of mathematical reasoning, we argue that both symbolic and nonsymbolic mathematical processes are grounded.

There is some evidence that intramathematical physical context influences success with arithmetic computations. Campbell (1994, 1999) explored error patterns in simple, single-operation computations with small numbers (0–9). Campbell found that error patterns were systematically affected by notation format (e.g., 9 + 0 vs. "nine plus zero") and argued that memory for simple mathematical operations is subserved not by an amodal, domain-general mathematical format, but by computation traces that record notational format. McNeil and Alibali (2004, 2005) similarly found

David Landy and Robert Goldstone, Department of Computer Science, Indiana University.

This research was funded by Department of Education, Institute of Education Sciences Grant R305H050116, and by National Science Foundation ROLE Grant 0527920. We would also like to gratefully acknowledge Adam Sheya, Steve Hockema, Ji Son, Art Glenberg, Erik Altmann, and Kevin Dunbar for their helpful comments and insight.

Correspondence regarding this article should be addressed to David Landy, Department of Computer Science, Indiana University, Bloomington, IN, 47405. E-mail: dlandy@cs.indiana.edu

that notational similarities between equations used to denote computations (3 + 4 + 5 = ?) affected the ability of prealgebra students to understand the use of equations to denote balance (3 + 4 = 5 + ?), and that the visual form of the equations strongly affected their perceived structure. Here, we explore how physical layout affects the segmentation of simple equations.

Segmenting (parsing) a notational form into its formal components is a difficult and routine part of mathematical reasoning. Current accounts of mathematical reasoning assume that this segmentation is cognitively executed through the application of formal rules to individual notational symbols (Anderson, 2005; Chandrasekaran, 2005; Koedinger & MacLaren, 2002; Stenning, 2002). These rules apply to abstract representations of the written expressions, which contain information regarding only the identity of the individual symbols and their sequence in the equation. Such accounts make the appealing and tempting assumption that the cognitive parser can extract abstract symbol sequences from physical notations in a trivial manner. However, dividing a visually presented notational expression into units appropriate for mathematical analysis may be a more complex and important process than it appears. In fact, a reasoner's syntactic interpretation may be influenced by notational factors that do not appear in formal mathematical treatments. In a model that translates atomic symbols into amodal logical symbols and combines them through syntactical rules, errors are expected to arise from three sources: misreading basic symbols, failing to transform the stored representation, or failing to generate an appropriate response. If, however, mathematical syntax is grounded in the visual format of notational displays, then manipulating the way terms in an equation are grouped visually should impact the effective order of precedence judgments. Some perceptual situations will support the mathematically correct grouping, whereas others will interfere with and weaken the mathematical representation by presenting a competing way to visually structure the equation.

In the four experiments described here, we evaluate whether these structural evaluations are sensitive to nonformal information available in visual displays, and whether and how that information is integrated into mathematical reasoning. In mathematical formalisms, nonordinal spatial proximity tends to indicate grouping. For instance, when people write expressions, they tend to put less space around multiplications than around additions, and they put even more space around equality signs (Landy & Goldstone, in press). This arrangement means that physical structures are often isomorphic to the formally sanctioned mathematical structure. Many fonts and typesetting programs also place less space around times signs than around plus signs. Parentheses, which formally indicate grouping, also create a regular visual segmentation. We hypothesized that people might use this typically available nonformal information to make grouping judgments, integrating this information with formal properties.

In the following experiments, participants were asked to judge whether individual simple equations were mathematically valid. Both valid and invalid equations were presented, and that validity was either sensitive or insensitive to the correct order of operations. In each experiment, a formally irrelevant grouping pressure was constructed to be consistent or inconsistent with the order of operations in the formal stimulus. Before turning to the experiments, we briefly define key terms.

## Key Terms

### *Order of Operations*

The natural language expression "five plus two times six" could be interpreted as resolving to either 42 or 17. The notational expression "5 + 2*6" is unambiguously equal to 17, because we agree to evaluate multiplications before additions, using parentheses to indicate the alternate interpretation. There is nothing mathematically necessary about multiplication that mandates that it be evaluated before addition. It is only within the context of standard algebraic notation that multiplication precedes addition. The rule that multiplication precedes addition is called the *order of operations*.

### *Permutation*

Tables 1 and 2 present examples of the equation structures used in these experiments. (In each table, a, b, c, and d represent randomly selected letters; different letters were used in each equation.) Because the same four variables appear on each side of an equation, and the same operators appear in the same order, whether or not the two expressions are necessarily equal depends only on the relative order of the operands on each side. These relative orderings are called *permutations*. There are 24 such permutations; in order to balance outcomes across the relevant conditions, only 8 were used. The first listed permutation is the identity permuta-

Table 1
*Experiment 1: Permutations and Mathematical Properties of Right-Hand Side Orderings, for the Operator Structure Plus-Times-Plus*

| Permutation | Possible left-hand side | Right-hand side | Valid | Valid if + precedes *? | Sensitivity |
|---|---|---|---|---|---|
| a b c d | a + b * c + d | = a + b * c + d | True | True | Insensitive |
| d c b a | a + b * c + d | = d + c * b + a | True | True | Insensitive |
| b c a d | a + b * c + d | = b + c * a + d | False | False | Insensitive |
| c a d b | a + b * c + d | = c + a * d + b | False | False | Insensitive |
| a c b d | a + b * c + d | = a + c * b + d | True | False | Sensitive |
| d b c a | a + b * c + d | = d + b * c + a | True | False | Sensitive |
| c d a b | a + b * c + d | = c + d * a + b | False | True | Sensitive |
| b a d c | a + b * c + d | = b + a * d + c | False | True | Sensitive |

*Note.* Stimuli were created by constructing a symbolic expression as the left-hand side of an equation and permuting its operands to construct the right-hand side.

Table 2
*Experiment 1: Permutations and Mathematical Properties of Right-Hand Side Orderings, for the Operator Structure Times-Plus-Times*

| Permutation | Possible left-hand side | Right-hand side | Valid | Valid if + precedes *? | Sensitivity |
|---|---|---|---|---|---|
| a b c d | a * b + c * d | = a * b + c * d | True | True | Insensitive |
| d c b a | a * b + c * d | = d * c + b * a | True | True | Insensitive |
| b c a d | a * b + c * d | = b * c + a * d | False | False | Insensitive |
| c a d b | a * b + c * d | = c * a + d * b | False | False | Insensitive |
| a c b d | a * b + c * d | = a * c + b * d | False | True | Sensitive |
| d b c a | a * b + c * d | = d * b + c * a | False | True | Sensitive |
| c d a b | a * b + c * d | = c * d + a * b | True | False | Sensitive |
| b a d c | a * b + c * d | = b * a + d * c | True | False | Sensitive |

tion, so the first row of Table 1 indicates that if the left-hand side of a particular stimulus were a + b*c + d, the right-hand side would be identical to the left.

## Validity

An equation is *valid* if its two sides are necessarily equal. The equation $5 + x = 6 + x - 1$ is valid because the expressions are equal regardless of the value of $x$, but $5*x = 6*x$ is not valid because it will not hold for all possible values of $x$. In Tables 1 and 2, the fourth column indicates whether each permutation was formally valid. Four of the permutations yield valid equations, and four yield invalid equations, so half of all presented equations were valid.

## Sensitivity

These experiments were designed to encourage participants to employ an incorrect addition-before-multiplication rule. Column 5 of Tables 1 and 2 indicates whether a particular permutation generates equations that are "valid" according to this incorrect operator order. When the equation's validity under the multiplication-before-addition rule differs from that found using the incorrect addition-before-multiplication rule (column 5), that equation is called *sensitive*. In Tables 1 and 2, a permutation is sensitive (column 6) only when the two validity columns have different values—when one entry is valid and the other invalid. Inspection of the table shows that four of the permutations are sensitive and four insensitive, and that sensitivity and validity are balanced.

Tables 1 and 2 differ only in operator structure. All expressions were either of the form a + b * c + d or of the form a * b + c * d. One result of this structure was that half of all equations in each of the four categories (valid/invalid crossed with sensitive/insensitive) had an addition as the leftmost operator and half had a multiplication as the leftmost operator, eliminating the possibility that our intended perceptual pressures might be confounded with a possible bias to bind operands in a left-to-right order.

## Consistency

In addition to variation in the formal mathematical properties, stimuli also varied in their physical layout. Our intention was to produce three classes of physical layouts. In *consistent* equations, a formally irrelevant physical feature is manipulated, creating a visual grouping that matches the formal grouping. In *inconsistent* equations, the same irrelevant feature creates a grouping around addition operations. Finally, *neutral* equations lack overt grouping pressures. See Table 3 for examples of each of the three types, which were constructed using physical proximity as the grouping pressure. Because these manipulations are entirely confined to the physical layout of the equation, they are independent of the mathematical variations defined above. Each of the mathematical variations was presented in each consistency mode.

Table 3
*Experiment 1: Samples of Stimuli Used*

| Permutation | Structure | Consistency | Validity | Example |
|---|---|---|---|---|
| a c b d | +*+ | Consistent | Valid | h + q*t + n = h + t*q + n |
| c a d b | +*+ | Consistent | Invalid | u + p*k + x = k + u*x + p |
| b a d c | *+* | Consistent | Valid | g*m + r*w = m*g + w*r |
| b c a d | *+* | Consistent | Invalid | y*s + f*z = s*f + y*z |
| d b c a | +*+ | Neutral | Valid | b + h * v + k = k + h * v + b |
| c d a b | +*+ | Neutral | Invalid | t + j * n + e = n + e * t + j |
| d c b a | *+* | Neutral | Valid | q * r + h * c = c * h + r * q |
| d b c a | *+* | Neutral | Invalid | w * g + x * j = j * g + x * w |
| a b c d | +*+ | Inconsistent | Valid | t+p * m+f = t+p * m+f |
| c d a b | +*+ | Inconsistent | Invalid | w+n * r+k = r+k * w+n |
| b a d c | *+* | Inconsistent | Valid | s * n+e * c = n * s+c * e |
| c a d b | *+* | Inconsistent | Invalid | g * a+w * j = w * g+j * a |

## Experiment 1

If algebraic reasoning uses visual groups as formal groups, then induced groupings ought to systematically influence the order of operations employed on the validity task. Proximity is a strong determiner of perceptual grouping (Koffka, 1935). In Experiment 1, by instructing participants to judge the validity of equations with nonstandard spatial relationships, we explored whether perceptual grouping affects mathematical reasoning.

There are many ways to fail on even a simple algebraic task such as this one. Solving these problems requires that participants match terms on the left side of the equation with terms on the right side; because each term must be checked, participants must also remember which terms have already been checked, and not reevaluate those. And, of course, participants must respect the relational structure of the equation, because in all of our test stimuli the same individual symbols appear on both sides of the equation. All of these task components are nontrivial, all are expected to cause some errors, and importantly, all are equally difficult for sensitive and insensitive equations. Thus, if physical spacing interferes with any of these aspects of the task (for instance by interfering with participants' ability to remember which symbols they have already checked), then spacing should affect both insensitive and sensitive equations. Our prediction, however, was that spacing would affect only accuracy on sensitive equations. If so, then that implies that spacing selectively affects order of operations judgments. The inclusion of insensitive equations thus helps clarify the role visual groups play in mathematical reasoning in this task.

Because consistent equations have, by our hypothesis, redundant information specifying syntax, such expressions might well be solved more quickly than neutral equations; similarly, one might expect inconsistent stimuli to take more time to correctly solve. However, by our hypothesis, the same processes are involved in solving sensitive and insensitive equations, thus consistency should affect response time on both types of problem, if on either. Therefore, if inconsistent equations take longer than consistent ones, they should do so on both insensitive and sensitive equations.

### *Method*

*Participants.* Fifty-one undergraduates participated in the experiment, which fulfilled a partial course requirement.

*Apparatus.* All expressions were presented in black text on a white background, using the Lucida Grande font on Macintosh computers. Monitor resolution was 1,024 × 768, and the monitor size was 38 cm. Participants sat approximately 55 cm from the monitors. The symbols were 3 mm across. Symbols were separated by 1.6 mm, 4.8 mm, or 12.7 mm, depending on condition (described in detail below). Therefore, the spaces between symbols subtended 0.17°, 0.5°, or 1.3° of visual arc in the narrow, neutral, and wide spacing cases, respectively. Participants used the keyboard to report validity judgments. The *P* and *Q* keys signified valid and invalid judgments, respectively.

*Design.* Our experiment was designed to orthogonally manipulate three factors: validity (the equation was valid or invalid), consistency (perceptual grouping was consistent, inconsistent, or neutral with respect to the normative multiplication-before-addition precedence rule), and sensitivity (whether or not validity/invalidity would be preserved given precedence of either multiplication-before-addition or addition-before-multiplication). We expected consistent equations to facilitate application of the correct multiplication-before-addition operator rule and inconsistent equations to promote application of an erroneous addition-before-multiplication rule.

Each participant viewed 240 test stimuli and 60 distractors. An individual stimulus consisted of a single symbolic equation. A response was a judgment of that equation's validity. Each stimulus equation consisted of two expressions (a left-hand side and a right-hand side) separated by an equals sign. Each expression contained four symbols, connected by three operators. Although operators appeared in the same order on both sides of the equation, the operand order could differ on the left- and right-hand sides. These constraints held for all test equations.

Each equation contained four unique symbols; due to their similarity to other symbols, the letters i, l, and o were omitted from the set of available letters. The 240 test equations were preceded by 10 unrelated hand-generated equations, and another unrelated equation appeared after every 5th test equation. Performance on these distractor equations was not included in any analysis. Distractors included different symbols on each side of the equation, division and subtraction, parentheses, and other complicated structures. The purpose of distractor equations was to discourage participants from solving problems using ad hoc shorthands or tricks based on the particular permutations, operator structures, and symbol constraints used in test equations.

*Procedure.* Participants were asked to proceed quickly, without sacrificing accuracy; instructions also reminded participants of the order of operations and stepped through a sample arithmetic computation. There was no time restriction; equations remained visible until participants responded. Immediately after the response, either a check mark or an X mark was displayed, depending on whether the response was correct or incorrect. This feedback remained on the screen for 300 ms, followed by a 700-ms delay, during which the screen was blank, and after which the next equation was displayed. Every participant received the equation set in a unique random order; the only constraint was that the first 10 equations and every 5th subsequent equation were randomly selected distractors. Participants received breaks after every 50 stimuli.

### *Results*

Seven participants whose neutral-trial performance was lower than 60% in the neutral spacing condition were removed from subsequent analysis, leaving 44 whose results were analyzed. A cutoff was employed in order to eliminate participants who lacked knowledge of the correct rule, or who simply did not attempt the task. Sixty percent was chosen as a cutoff because of a natural division in the participant's scores. In all four experiments discussed here, the reported effects remain significant when all data are included in the analysis.

Accuracy differed between the sensitivity and consistency factors according to a two-way within-participants analysis of variance (ANOVA). Both main effects were significant. Participants were less accurate on sensitive judgments, $F(1, 42) = 54.3$, $MSE = 7.56$, $p < .001$. Accuracy was highest for spatially consistent equations, intermediate for neutral, and lowest for inconsistent equations, $F(2, 41) = 62.621$, $MSE = 3.789$, $p < .001$. For our purposes, however, the most interesting effect is the interaction: participants were far more affected by spatial consistency when the order of operations judgment affected the correct answer, $F(2, 41) = 55.7$, $MSE = 3.1$, $p < .001$. For insensitive trials, the mean difference between consistent and inconsistent accuracies was only 2%; for sensitive trials, the difference rose to 38.7%. Mean accuracies for each factor are shown in Figure 1.

The data were also analyzed with normative validity as a factor in the analysis. Accuracy was higher on valid (86.4%) than on invalid (82.2%) equations, $F(1, 42) = 52.93$, $MSE = 5.1$, $p < .001$. Validity did not interact with either consistency or sensitivity. Inclusion of validity did not affect the main results of interest in this or any subsequent experiments, and will not be discussed.

Another useful measure of the impact of the consistency manipulation is the fraction of participants who were more accurate on consistent trials than on inconsistent trials, in the sensitive and insensitive conditions. On insensitive trials, 25 of the 44 participants had higher accuracies on consistent trials, 16 had higher accuracies on inconsistent trials, and 3 had equal accuracies on each. If consistency had no effect on accuracy, then we would expect about as many participants to have higher scores in the consistent condition than in the inconsistent condition. According to the binomial test of significance, and discounting the 3 participants whose scores were equal in both conditions, the fraction of participants who scored higher in the consistent condition on insensitive trials (25 of 41) is consistent with a probability of 0.5, and with no significant effect of consistent versus inconsistent trials. Only 1 participant showed the reverse; this proportion is significantly different from chance ($p < .0001$) and also from the proportion of participants showing a benefit of consistency on insensitive trials ($\chi^2 = 15.7$, $p < .001$). These results (along with the results of a binomial by-participants analysis from the remaining three experiments) are summarized in Table 4.

While the principal theoretical interest is in the existence of an impact of consistency on accuracy, response time (RT) differences across conditions were large and robust. Mean RT on all (accurately judged) trials was 2,264.48 ms, with a standard error of 977.76 ms. According to a two-way within-subjects ANOVA, RT differed significantly across consistency conditions, $F(1, 42) = 8.8$, $MSE = 715{,}734$, $p = .005$ (see Figure 2). There was no significant main effect of sensitivity on RT, nor was there an interaction between sensitivity and consistency; this is strongly consistent with the perceptual grouping account but is not necessarily inconsistent with a more amodal symbolic account.

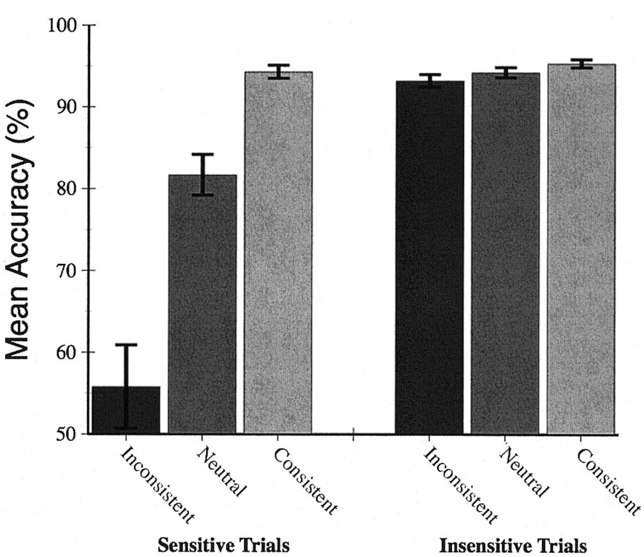


*Figure 1.* Experiment 1: Spacing. Values are mean accuracies, divided across consistency and sensitivity conditions, averaged across subjects. Error bars are standard errors.

### *Discussion*

The physical spacing of formal equations has a large impact on successful evaluations of validity. Furthermore, the impact seems to be limited to order of operations evaluations. Symbols placed physically closer together tend to be evaluated as syntactically bound. It is not immediately obvious that interference and facilitation of the specific mathematical order of operations is responsible for the effects seen in Experiment 1. For example, one might suppose that because the unusually spaced equations are nonstandard, it would be more difficult to treat them as mathematical forms or to extract relevant information from them. However, if participants were simply unable to apply symbolic processes learned with neutrally spaced equations to unusually spaced stimuli, and were consequently making nonmathematical judgments on these displays, then we would expect performance to be impaired on both inconsistent and consistent stimuli, because both were spaced very unusually. (It is quite likely that naturally encountered equations tend to be slightly consistently spaced. However, in the equations we used, the spacings were quite exaggerated). Instead, performance was higher on consistent stimuli than on the presumably more familiar neutral cases.

A further indication that this benefit really is mathematical in character is that both facilitation and interference occurred only for sensitive permutations of the symbols. Although performance on insensitive equations was higher than on neutral sensitive, it was not at ceiling; although accuracy on some insensitive permutations (such as the identity permutation) was very high, the mean accuracy on consistent insensitive equations was not substantially higher than on consistent sensitive equations (94.8% vs. 93.7%, respectively; $p \sim .17$ by a within-subjects $t$ test). The relative sparing of equations insensitive to the order of operations strongly indicates that participants were in fact engaging their mathematical faculties by using the grouping implied by the visual structure, and not simply failing to treat the unusually spaced terms as equations.

An interesting possibility is that different participants may have used different strategies when approaching the validity task; because of their complexity, mathematical tasks such as validity are often amenable to a variety of different solution strategies. If individual participants pursued different strategies, one might expect that some strategies would be highly sensitive to visual form, whereas other strategies—perhaps linguistic or symbolic —would be more-or-less unaffected by a perceptual manipulation. However, if this were the case, we would expect that many participants—those pursuing linguistic strategies, for instance—would be largely unaffected by spacing. Our results are incompatible with this expectation: only 1 of the 44 participants correctly solved more inconsistent trials than consistent trials in the sensitive con-

Table 4
*Experiments 1–4: No. of Participants Whose Accuracy on Consistent Trials Was Higher Than, Lower Than, or Equal to Their Accuracy on Inconsistent Trials*

| | Trial type with highest accuracy | | | |
|---|---|---|---|---|
| Condition | Consistent | Inconsistent | Equal | Binomial significance (null hypothesis $p = 0.5$) |
| Exp. 1: Proximity | | | | |
| Insensitive trials | 25 | 16 | 3 | ~.22 |
| Sensitive trials | 43 | 1 | 0 | <.001 |
| Exp. 2: Region | | | | |
| Insensitive trials | 17 | 12 | 16 | ~.46 |
| Sensitive trials | 35 | 5 | 5 | <.001 |
| Exp. 3: Similarity | | | | |
| Insensitive trials | 13 | 17 | 3 | ~0.52 |
| Sensitive trials | 18 | 5 | 7 | <.01 |
| Exp. 4: Familiarity | | | | |
| Insensitive trials | 30 | 22 | 13 | ~.33 |
| Sensitive trials | 36 | 14 | 15 | <.01 |

dition. The impact of proximity on performance seems to be ubiquitous throughout our sample.

This impact is not predicted by existing accounts of notational mathematical reasoning. It is quite possible that an account that retains the notion of mathematical knowledge as knowledge of a formal syntax could be made to predict the systematic influence we observed, but such an account would not provide a theoretically motivated explanation of why proximity and order of operations are connected in this way. Because this connection is not obviously motivated by either overt instruction or mathematical rules, it is not clear why such a connection would appear in a system driven by structural rules between singular terms.

Other potential accounts cannot be eliminated by the data in Experiment 1. One important possibility is that experience with parenthesized equations interferes with participants' ability to treat space as unimportant. This could result from an (implicit or explicit) inference that the spacing information must be relevant because it is so noticeably present, and therefore that it must signify parentheses; it could also result from the perceptual similarity of the inconsistently spaced equations to familiar parenthesized ones. One straightforward way to disentangle the parenthesis and the perceptual grouping explanations is to construct equations with grouping pressures that do not resemble parenthesized expressions.

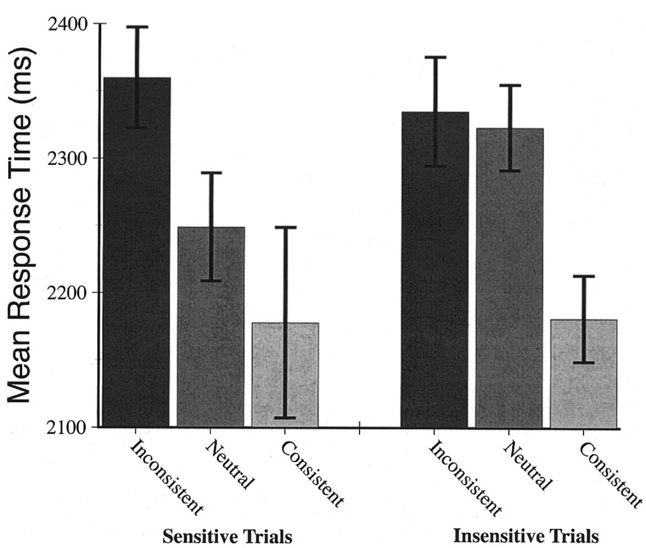


*Figure 2.* Experiment 1: Spacing. Values are mean response times (in milliseconds) divided across consistency and sensitivity conditions. Error bars are standard errors.

Another possible explanation of the connection between proximity and order of operations is that participants use their knowledge of written language in interpreting mathematical forms. In written English, spaces are used to specify the appropriate grouping of letters into words; participants might implicitly apply training in visual word segmentation to equation interpretation. However, in English, spacing in particular, and not grouping in general, is used to indicate segments. Experiment 1 confounds associations that are specifically tied to physical spacing with our hypothesis that perceptual grouping in general contributes to mathematical segmentation and that proximity cues grouping. In the remaining experiments, we confirm the robustness of the hypothesized relationship between perceptual grouping and relational interpretation in mathematics notation by exploring the impact of other classic grouping principles on notational reasoning.

## Experiment 2

This experiment used the common region grouping principle to encourage subjects to make particular visual interpretations. According to the common region principle, visual elements that are located within the same visual region are grouped (Palmer, 1992). Regions can be demarcated using color shifts, visual separators, or object boundaries. In Experiment 2, visual regions were created around subexpressions to selectively group those expressions.

### Method

*Participants, apparatus, and procedure.* Forty-seven Indiana University undergraduates participated in exchange for partial course credit. Apparatus and procedure were identical to those used in Experiment 1.

*Design.* The design of Experiment 2 was nearly identical to that of Experiment 1. Once again, participants judged the validity of visually presented symbolic equations. The mathematical variations in the stimuli were identical: Each expression consisted of four terms, separated by either the plus-times-plus or the times-plus-times operator structure. As in Experiment 1, the operands and operators on the right- and left-hand sides of the equation were identical; only the order of the operands differed. The eight permutations used in Experiment 1 were used again in Experiment 2. Once again, participants viewed 240 test equations and 60 distractor equations.

The sole difference between Experiments 1 and 2 was the perceptual manipulation employed. In Experiment 2, the actual mathematical forms were not altered (nor was anything in the physical line of the equation). Instead, embedding some of the symbols inside implied oval-shaped regions created perceptual structures. Of the 240 test stimuli, 160 contained ovals around some of the symbols (see Figure 3). In 80 of the stimuli, the perceptual groups were designed to be consistent with the order of operations: Multiplicands were placed together in a common region (along with the multiplication operator). In the other 80, the perceptual groups were inconsistent with the order of operations. In these, addends (and addition operators) were embedded in a common region. The remaining 80 test equations lacked ovals and thus were considered neutral. The distractor equations were similar to the neutral test equations but were more general; symbols sometimes differed between the left- and right-hand expressions, and operators appeared in varied orders.

## Results

Participants who scored less than 60% over all trials were removed from subsequent analysis. Seven participants were re-

| Permutation | Structure | Consistency | Validity | Example |
|---|---|---|---|---|
| a c b d | +*+ | Consistent | Valid | h + q * t + n = h + t * q + n |
| c a d b | +*+ | Consistent | Invalid | u + p * k + x = k + u * x + p |
| b a d c | *+* | Consistent | Valid | g * m + r * w = m * g + w * r |
| b c a d | *+* | Consistent | Invalid | y * s + f * z = s * f + y * z |
| d b c a | +*+ | Neutral | Valid | b + h * v + k = k + h * v + b |
| c d a b | +*+ | Neutral | Invalid | t + j * n + e = n + e * t + j |
| d c b a | *+* | Neutral | Valid | q * r + h * c = c * h + r * q |
| d b c a | *+* | Neutral | Invalid | w * g + x * j = j * g + x * w |
| a b c d | +*+ | Inconsistent | Valid | t + p * u + f = t + p * u + f |
| c d a b | +*+ | Inconsistent | Invalid | w + n * r + k = r + k * w + n |
| b a d c | *+* | Inconsistent | Valid | s * n + e * c = n * s + c * e |
| c a d b | *+* | Inconsistent | Invalid | g * a + w * j = w * g + j * a |

*Figure 3.* Experiment 3: Examples of stimuli used. Implied oval-shaped regions embedded in the equations were used to create perceptual grouping.

moved for this reason, leaving 40 participants whose results were analyzed.

Accuracy was affected by the sensitivity and consistency factors according to a two-way, within-participants ANOVA. Both main effects were significant. Participants were less accurate on sensitive judgments than on insensitive judgments, $F(1, 38) = 34.7$, $MSE = 36.1$, $p < .001$. On sensitive trials, accuracy was highest for spatially consistent equations, intermediate for neutral, and lowest for inconsistent equations, $F(1, 38) = 23.6$, $MSE = 17.8$, $p < .001$. The interaction term was also significant. Participants were far more affected by spatial consistency on sensitive trials than on insensitive ones, $F(1, 38) = 25.8$, $MSE = 17.0$, $p < .001$. Mean accuracies for each condition are shown in Figure 4.

Once again, we analyzed the performance of individual participants by comparing accuracy on consistent trials with accuracy on inconsistent trials by means of a binomial significance test. The results of this analysis are summarized in Table 4. In this experiment, no response time results reached significance.

### Discussion

It is clear that the manipulation of common region has a strong impact on validity judgments. Because the manipulation affects accuracy only on permutations that are sensitive to the order of operations, it seems that perceptual grouping directly affects validity judgments. This result is expected only if visual processes are directly applied to multisymbol units to perform validity judgments. The results of Experiment 2 provide support for our interpretation of the previous experiment: that perceptual grouping, and not literal similarity to parenthesized equations, is responsible for the consistency effect. However, the ovals used to determine common region are highly salient features and (to some observers at least) resemble parentheses; it is quite possible that participants (contrary to instructions and trial-by-trial feedback) treated the illusory ovals as a task demand to group against the normal order of operations. In the remaining two experiments, more subtle manipulations of equation format were employed to provide converging support for the perceptual grouping interpretation.

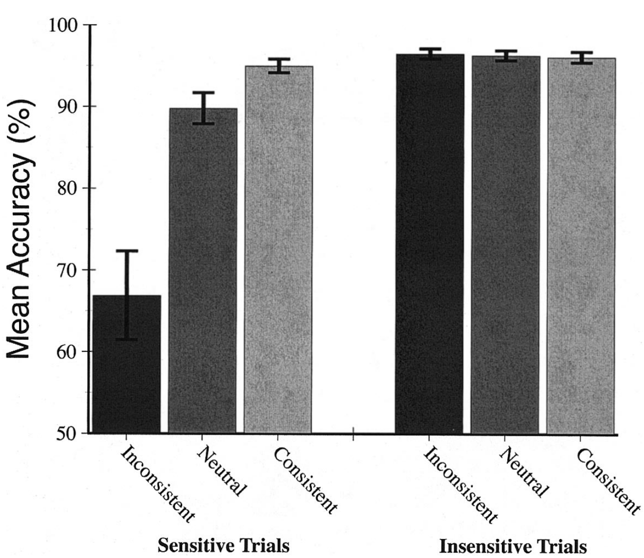


*Figure 4.* Experiment 2: Common region. Values are mean accuracies, divided across consistency and sensitivity conditions, averaged across subjects. Error bars are standard errors.

## Experiment 3

Experiment 3 examined a kind of pressure different from that of the previous experiment: perceptual similarity. Here, rather than using single letters or numbers as the symbols to be grouped, we used compound parenthesized expressions. These expressions came in one of several regular structures. Because objects with similar shapes tend, in general, to group together visually (Koffka, 1935), we hypothesized that identical mathematical and visual structure would cause a similar grouping pressure, which would then assist or impair application of the order of operations.

### Method

*Participants, apparatus, and procedure.* Thirty-three Indiana University undergraduates participated in exchange for fulfillment of a course requirement. Apparatus and procedure were identical to those used in Experiment 1.

*Design.* The design was similar to that of Experiment 1. The same basic mathematical setup was employed: Each stimulus consisted of an equation. Each expression in the equation consisted of four terms, separated by either the plus-times-plus or the times-plus-times operator structure. Again, the same operands and operators appeared on both sides of an equation. Once again, the only difference between the left- and right-hand sides was the order in which the operands appeared. The same 8 permutations used in Experiments 1 and 2 were also used for Experiment 3, so the design employed the same sensitive and insensitive categories, and once again counterbalanced validity and operator order within these categories. Therefore, there were again 240 test equations, with 60 distractors. Once again, participants were instructed to determine the validity of a putative equation and to press a corresponding key.

The sole difference between Experiments 1 and 2 and Experiment 3 is the nature of the consistency manipulation. Whereas physical spacing was manipulated in Experiment 1, internal relations between the four operands were varied in Experiment 3. The operands in Experiment 3 were compound parenthesized terms such as $(8/m)$ rather than simple letters or numbers. This feature facilitated the manipulation of internal operand similarity through the systematic manipulation of structure; for instance, $(8/m)$ is more similar to $(2/q)$ than to $(y*y*y)$. There were four basic term schemas (see Table 5); each actual operand term was derived from one of these types by replacing each of the slots (the words *letter* and *number*) in the schema with a randomly chosen selection of letters and numbers. Physical similarity is assumed to be high among pairs of terms derived from a common schema, and low between terms derived from different schemas. Once again, stimuli were constructed so as to be perceptually consistent, inconsistent, or neutral with respect to the mathematical order of operations. With inconsistent equations, highly similar terms—terms derived from a common schema—surrounded multiplication operators, while dissimilar terms surrounded addition operators. With inconsistent equations, terms surrounding additions were derived from common schemas, while terms surrounding multiplications were

Table 5
*Operand Schemas Used in Experiment 3*

| Index | Schema |
|---|---|
| 1 | (letter / number) |
| 2 | (letter * letter * letter) |
| 3 | (letter + number * letter) |
| 4 | (number letter − letter) |

*Note.* The words *letter* and *number* indicate slots in the schema. In actual stimuli, the words would be replaced by actual (randomly selected) letters and numbers. Spaces are added to the schema for clarity; in actual stimuli, the operands contained no spaces.

structurally similar. On neutral trials, all four terms were drawn from the same schema, and this was always Schema 1. The use of four identical structures, rather than a random disjoint combination, prevented contamination from any implicit similarities between the different structures. Sample stimuli are presented in Table 6.

One unfortunate feature of the manipulation used here is that it can be presented only on one side of an equation. For example, consider a consistent equation with operator order times-plus-times, in which the first two terms are drawn from Schema 2, and the last two from Schema 3. The left-hand side might be $(5a - x) * (3n - j) + (2/g) * (8/y)$. This is a consistent expression because the multiplicands are drawn from a common schema, whereas the addends are drawn from different schemas. However, consider what happens when the right-hand side is constructed using the acbd permutation. The right-hand side will then be $(5a - x) * (2/g) + (3n - j) * (8/y)$, which is neither consistent nor inconsistent. This complication is inherent in the basic experimental design and afflicts most of the permutations used in the experiment, but it is not fatal. Because the consistency structure does exist on the left-hand side of the equation (which is usually read before the right-hand side), and no strong conflicting structure exists on the right, any systematic effects of the manipulation are still attributable to the consistency structure of the operand terms on the left-hand side. In all cases, expressions are genuinely valid or invalid, and therefore participants are given a sensible task.

## Results

Three participants were removed from the analysis because they did not meet a criterion of 60% performance in the neutral trials. Once again, a two-way repeated measures ANOVA with accuracy as the dependent variable confirmed significant main effects of sensitivity, $F(1, 28) = 22.1$, $MSE = 3.12$, $p < .001$, and consistency, $F(1, 28) = 5.608$, $MSE = .067$, $p < .01$, and a significant interaction between sensitivity and consistency, $F(1, 28) = 4.96$, $p = .01$. The mean accuracies within each condition are displayed in Figure 5. Note that, once again, the increase in accuracy with consistent alignment is much stronger in equations that are sensitive to the order of operations than in those that are equal or unequal regardless of grouping. The number of individual participants with a higher consistent than inconsistent accuracy, along with binomial tests of significance, is summarized in Table 4. In this experiment, no response time results reached significance.

## Discussion

The results of Experiment 3 support those of Experiments 1 and 2. The similarity of effects of physical spacing and internal structure on order of operations evaluation implies that each affects that evaluation in a similar way. Generally, these results support the perceptual grouping account. Except for a common impact on grouping, the physical manipulation employed in Experiment 3 was quite different from that of the first two experiments. Unlike the results of Experiment 1, the results of manipulating similarity cannot be accounted for by physical similarity of the given equations to parenthesized ones. Furthermore, the manipulation is quite subtle and clearly semantic, making it unlikely that participants inferred parentheses.

The manipulation used here confounds two subtly different kinds of similarity. First, templates drawn from the same type are physically alike: They have parentheses, operators, and letter/numerals in the same physical locations. This makes two terms drawn from a common schema more perceptually similar than two drawn from different schemas. For instance, a physical template constructed from a Schema 1 term will better match another Schema 1 term than a term derived from Schema 2. The second kind of similarity is semantic: Two terms drawn from a common schema will share semantic structure; for instance, any term drawn from Schema 1 comprises an addition of an atomic symbol and a product. We suspect that individual terms were not processed in sufficient depth to make the analogical structural commonalities important sources of similarity. Problems could always be solved

Table 6
*Experiment 3: Samples of Stimuli Used*

| Permutation | Consistency | Validity | Sample stimulus |
|---|---|---|---|
| a c b d | Consistent | Valid | (d*d*d) + (4/q) * (9/m) + (p+2*x) = (d*d*d) + (9/m) * (4/q) + (p+2*x) |
| b a d c | Consistent | Valid | (t*t*t) * (m*m*m) + (5/h) * (2/v) = (m*m*m) * (t*t*t) + (2/v) * (5/h) |
| c a d b | Consistent | Invalid | (8/g) + (6a−n) * (3m−p) + (a*a*a) = (3m−p) + (8/g) * (a*a*a) + (6a−n) |
| b c a d | Consistent | Invalid | (f*f*f) * (u*u*u) + (y+3*k) * (a+2*q) = (u*u*u) * (y+3*k) + (f*f*f) * (a+2*q) |
| d b c a | Neutral | Valid | (g+4*v) + (u+1*b) * (c+9*f) + (r+8*v) = (r+8*v) + (u+1*b) * (c+9*f) + (g+4*v) |
| d c b a | Neutral | Valid | (t+6*h) * (y+2*q) + (a+5*x) * (k+9*r) = (k+9*r) * (a+5*x) + (y+2*q) * (t+6*h) |
| c d a b | Neutral | Invalid | (j+7*t) + (u+1*b) * (w+9*y) + (r+8*v) = (w+9*y) + (r+8*v) * (j+7*t) + (u+1*b) |
| a c b d | Neutral | Invalid | (t+6*h) * (y+2*q) + (a+5*x) * (k+9*r) = (t+6*h) * (a+5*x) + (y+2*q) * (k+9*r) |
| a b c d | Inconsistent | Valid | (8/h) + (3/n) * (w+4*c) + (h+1*b) = (8/h) + (3/n) * (w+4*c) + (h+1*b) |
| b a d c | Inconsistent | Valid | (3/b) * (j*j*J) + (z*z*z) * (7u−y) = (j*j*j) * (3/b) + (7u−y) * (z*z*z) |
| c d a b | Inconsistent | Invalid | (d*d*d) + (q*q*q) * (4/q) + (9/m) = (4/q) + (9/m) * (d*d*d) + (q*q*q) |
| c a d b | Inconsistent | Invalid | (v*v*v) * (3r−h) + (8e−u) * (2/s) = (8e−u) * (v*v*v) + (2/s) * (3r−h) |

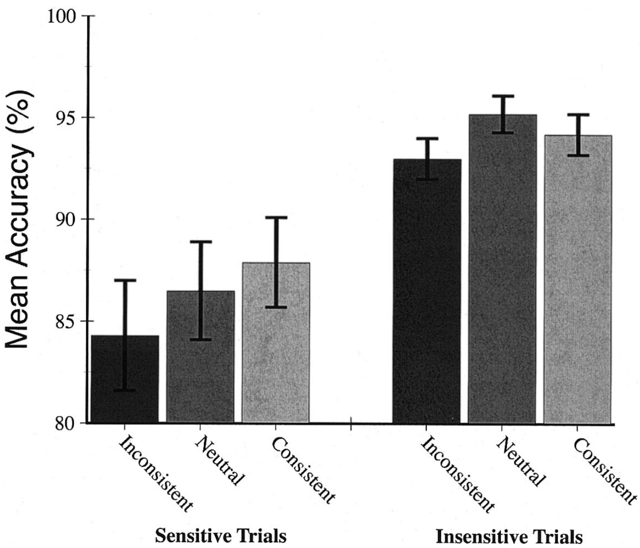


*Figure 5.* Experiment 3: Shape similarity. Values are mean accuracies, divided across consistency and sensitivity conditions, averaged across subjects. Error bars are standard errors.

without deep structural analysis of the parenthesized expressions. However, either source of similarity could in principle account for the effects of the consistency manipulation. Of course, neither manipulation should normatively affect validity interpretations or grouping, and either would be interesting in its own right. Our interpretation—that the perceptual structure is principally responsible—is only one possible interpretation, which we favor because it is consistent with the results of Experiment 1.

Also, unlike the results of the previous experiments, in Experiment 3 many individual participants had higher scores on inconsistent trials than on consistent trials. Because the effect is smaller, this result is not incompatible with the assumption that participants were all using the visual information; however, this result is also compatible with the possibility that participants may have used a variety of strategies on this experiment.

Experiment 4 reinforces Experiments 1, 2, and 3, and explores further whether the grouping processes of the first two experiments are driven only by early perceptual forces. Both proximity and similarity are early perceptual features. In Experiment 4, we manipulated the presence of a learned category that generates a late grouping pressure: alphabetic proximity.

## Experiment 4

The human visual system groups objects not just by virtue of their strictly perceptual properties, but also through learned associations and pattern familiarity. In Experiment 4, we evaluated the impact of a highly familiar pattern—the alphabet—on mathematical grouping. We explored whether alphanumeric proximity/similarity affects participants' effective order of operations. Although this visual grouping principle may be expected to be much weaker than spatial proximity, it is a useful principle to test because it does not create equations that are easily mistaken for parenthesized structures.

### Method

*Participants, apparatus, and procedure.* Seventy-two Indiana University undergraduates participated in the experiment in exchange for fulfillment of a course requirement. Apparatus and procedure were identical to those used in Experiment 1.

*Design.* The design was very similar to that of Experiment 1. Once again, participants judged the validity of a series of putative equations. The same operator structures and permutations were used, as were the three consistency classes, making again 48 total types of equation. Again, all participants judged 300 equations, with 240 randomly ordered test equations and 60 randomly ordered but evenly spaced distracter equations.

In Experiment 4, we manipulated consistency between structure and form by varying the alphabetic proximity of terms across the two operators. As in Experiment 1, the four operand terms were single letters or numbers. In the consistent condition, terms separated by a multiplication were adjacent in the alphabet, and in the inconsistent condition, terms separated by the addition were adjacent. The symbols picked were always from adjacent elements in the sets {a,b,c,d}, {p,q,r,s}, {w,x,y,z}, or {2,3,4,5}, which guaranteed that symbols which were not adjacent were widely separated alphabetically. All letters and numbers excluding i, l, and o were used in the neutral condition; however, the randomly selected letters were chosen so that none were separated by fewer than two other letters.

As in Experiment 3, the grouping pressures arise from properties intrinsic to the tokens, so it is impossible to preserve the same grouping pressures on the left- and right-hand sides of the equation. The equations are categorized by the grouping pressures hypothesized to exist on their left-hand sides. Sample stimuli are presented in Table 7.

### Results

Seven participants whose performance in the neutral condition was below 60% were removed from the analysis. A two-way within-participants ANOVA using accuracy as the dependent measure revealed main effects both of consistency, $F(1, 63) = 7.62$, $MSE = .015$, $p < .001$, and sensitivity, $F(1, 63) = 25.3$, $MSE = 0.64$, $p < .001$. There was also a significant interaction; the effect of consistency was larger on sensitive equations than on insensitive expressions, $F(1, 63) = 4.55$, $MSE = 0.007$, $p < .05$. The mean values along with standard errors are displayed in Figure 6. Counts of individual participant differences on consistent and inconsistent trials are summarized in Table 4. No results regarding response time reached significance.

### Discussion

Although the effect of consistency exhibited in Experiment 4 is much smaller than that of Experiments 1 and 2, it is worth taking a moment to be surprised that it is there at all. Although it is quite usual for people to prefer certain letters or letter sets for certain semantic roles in an algebraic expression (compare $h = tu + n$ with $y = mx + b$), it is counterintuitive (to us, at least) that general alphabetic proximity should affect simple mathematical judgments, and especially that it should apparently interact with the application of the order of operations so specifically. Although we

Table 7
*Experiment 4: Samples of Stimuli Used*

| Permutation | Operator structure | Consistency | Validity | Example |
|---|---|---|---|---|
| a c b d | +*+ | Consistent | Valid | a + p * q + z = a + q * p + z |
| c d a b | +*+ | Consistent | Invalid | u + 2 * 3 + e = 3 + e * u + 2 |
| b a d c | *+* | Consistent | Valid | k + m * n + t = m + k * t + n |
| b c a d | *+* | Consistent | Invalid | w * x + q * r = x * q + w * r |
| d b c a | +*+ | Neutral | Valid | a + u * j + m = m + u * j + a |
| c d a b | +*+ | Neutral | Invalid | r + w * q + b = q + b * r + w |
| d c b a | *+* | Neutral | Valid | r * f + w * n = n * w + f * r |
| a c b d | *+* | Neutral | Invalid | 4 * u + s * e = 4 * s + u * e |
| a b c d | +*+ | Inconsistent | Valid | r + s * b + c = r + s * b + c |
| c d a b | +*+ | Inconsistent | Invalid | a + b * 3 + 4 = 3 + 4 * a + b |
| b a d c | *+* | Inconsistent | Valid | 4 * b + p * r = b * 4 + r * p |
| c a d b | *+* | Inconsistent | Invalid | c * 4 + 5 * r = 5 * c + r * 4 |

cannot rule out the possibility that spacing, structure, and alphabetic proximity affect order of operations judgments for unrelated reasons, or for a reason other than their impacts on grouping, the factors known to influence perceptual grouping provide support for the hypothesis that mathematical grouping interacts with perceptual grouping generally.

Unlike the pressures manipulated in the previous experiments, pressure from alphabetic proximity is categorical rather than strictly perceptual. Furthermore, the pressure depends on a specific, highly learned pattern. This experiment therefore explored the effect of a late, association-based grouping pressure on algebraic order of operations.

## General Discussion

The current experiments show an influence of perceptual grouping on mathematical problem solving. The results are noteworthy for several reasons. First, they demonstrate a genuine cognitive illusion in the domain of mathematics. The criteria for cognitive illusions in reasoning are that people systematically show an influence of a factor in reasoning; that the factor should normatively not be used; and that people agree, when debriefed, that they were wrong to use the factor (Tversky & Kahneman, 1974). These criteria are met in our paradigm. Our participants were systematically influenced by the grouping variables of physical spacing, alphabetic familiarity of variable names, and notational form. The normative formalism of mathematics does not include these factors.

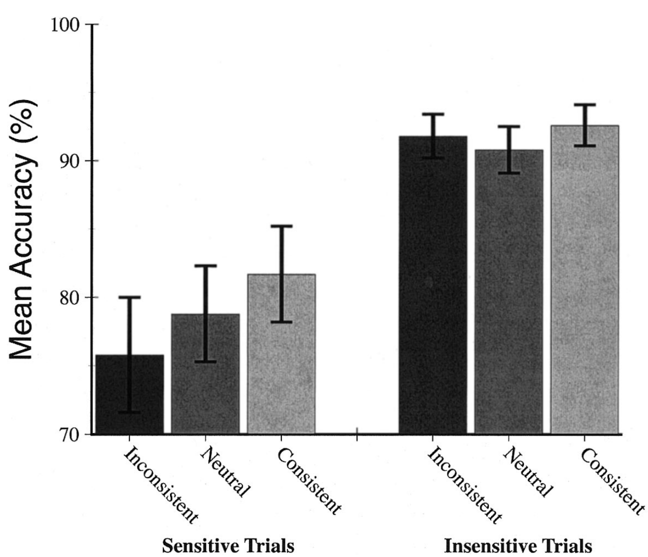


*Figure 6.* Experiment 4: Familiarity. Values are mean accuracies, divided across consistency and sensitivity conditions, averaged across subjects. Error bars are standard errors.

Postexperimental reviews indicated that some participants realized that that they were affected by grouping. Other participants believed that their responses were in accordance with the standard order of operations. In both cases, participants knew that responding on the basis of space, alphabetic familiarity, and similarity of notation were incorrect, but they continued to be influenced by these factors. The influence did decrease somewhat by the end of 240 trials, but even at the end of the experiment, the difference in accuracy between consistent and inconsistent perceptual grouping, for example using the spacing manipulation of Experiment 1, was 14%.

The second impressive aspect of the results is that participants continued to show large influences of grouping on equation verification even though they received trial-by-trial feedback. One might have argued that participants were influenced by grouping only because they believed that they could strategically use superficial grouping features as cues to mathematical parsing. However, constant feedback did not eliminate the influence of these superficial cues. This suggests that sensitivity to grouping is automatic or at least resistant to strategic, feedback-dependent control processes. Grouping continued to exert an influence even when participants realized, after considerable feedback, that it was likely to provide misleading cues to parsing. Physical spacing might be thought to be perceptually early and hence resistant to strategic control. However, alphabetic and format similarity are not computed early in the perceptual processing stream and yet still show robust, feedback-resistant influences.

The third impressive aspect of the results is that an influence of grouping was found in mathematical reasoning. Mathematical reasoning is often taken as a paradigmatic case of purely symbolic reasoning, much more so than language, which, in its spoken form, is produced and comprehended before children have acquired formal operations (Inhelder & Piaget, 1958). Some branches of mathematics, notably topology and geometry, are often argued to use visuospatial routines (Hadamard, 1949). However, algebra is, according to many people's intuition, the clearest case of widespread symbolic reasoning in all human cognition. To find an influence of variable names in algebraic calculation is striking. Participants know that the names given to variables are not supposed to influence calculations, but their behavior indicates exactly such an influence.

The current work is consistent with other literature showing that people may use perceptual cues instead of rules even when they know that the rules should be applied. For example, Allen and Brooks (1991) provided participants with an easy rule for categorizing cartoon animals into two groups. If the animal had at least two of the features long legs, angular body, or spots then it was a builder; otherwise it was a digger. Participants were trained repeatedly on eight animals. Then participants were transferred to new animals. Some of the animals looked very similar to one of the eight training stimuli but belonged to a different category. These animals were categorized more slowly and less accurately than animals that were equally similar to an old animal and belonged in the same category as the old animal. Participants seem not to have been able to ignore perceptual similarities between old and new animals, even though they knew a straightforward and perfectly accurate categorization rule. As another example, Palmeri (1997) had participants judge the numerosity of random patterns with 6–11 dots. After several days of training on a specific set of stimuli, the classification accuracy of new stimuli was heavily influenced by their perceptual similarity to old, trained patterns. A picture with 7 dots that was perceptually similar to one of the trained 6-dot patterns was often judged to have 6 dots. Participants knew the rule for counting but ended up being influenced by similarity factors that they knew to be potentially misleading. These and other (Keil, 1989; Ross, 1987; Smith & Sloman, 1994) studies suggest that people often use perceptual similarity as a basis for their judgments even when they know the correct formal rules that should be applied.

Our current studies extend these previous demonstrations in two important ways. First, we demonstrate intrusions of perceptual grouping even in algebraic reasoning tasks that are taken as archetypal cases of symbolic rule processing. Unlike the novel and unfamiliar rules used by Allen and Brooks (1991), the order of operations rules used in our study were familiar to participants, thus they might have been expected to deploy these rules without interference from mathematically irrelevant groupings. Second, we show influences of perceptual similarity not between items that affects whether one item will remind a person of another item, but also perceptual similarity within an item that leads to some elements of the item to being grouped together.

### *Possible Mechanisms Mediating Mathematical Grouping*

Both formal models (Butterworth, Zorzi, Girelli, & Jonckheere, 2001; Anderson, 2005) and more theoretical conceptions (Stenning, 2002; Chandrasekaran, 2005) conceive of notation interpretation as a two-stage process: First, a visual scene is decomposed into a string of basic graphemes. Second, abstract representations of these basic graphemes are recomposed using the internally represented laws of an algebraic grammar. The first process is a visual one: the perceptual system uses visible properties to categorize a symbol into one of a collection of discrete types. At the intermediate stage, representation of the size, shape, nonordinal position, color, and other concrete physical characteristics contained in actual physical symbol tokens is dropped, and the symbol information passed to the reconstruction mechanism is more-or-less abstract. Although these symbols may still legitimately be cast as visual (Koedinger & MacLaren, 2002), only the category of the symbol is retained for subsequent processing. The only relational structure that is retained is the order in which the symbols appear. We call such accounts *two-stage accounts*, because they assume that visual processes precede and are independent of the formal or cognitive processes that implement mathematical reasoning.

In the two-stage accounts conception, notational perception is cleanly separated from mathematical reasoning proper. For this reason, opportunities for an interaction between the two processes are strictly limited. Symbols that are poorly written may be difficult to correctly categorize; and deliberate interference with working memory might make it difficult to remember one symbol while reading the next. However, the decomposition of the visual scene into discrete adjacent symbols makes physical relations between the individual symbols necessarily irrelevant to mathematical relationships (Stenning, 2002; Chandrasekaran, 2005).

A parsimonious account of the relationship between validity behavior and perceptual grouping demonstrated here is that the reasoning processes do not initially represent a viewed equation as a sequence of individual notational graphemes. Instead, they may use the hierarchy of groups automatically constructed by the visual system as the representations over which reasoning processes occur, resorting to the analysis of individual graphemes only when necessary. We imagine reasoning as an interaction between typically visual processes and central reasoning. In the case of validity, a parsimonious model might go something like this: Perceptual processes group ("chunk") scenes at a particular level of the part–whole hierarchy, and present those chunks (or pointers to those visual chunks) to a subsequent reasoning process. That reasoning process verifies that the same chunks appear on the left- and right-hand sides of the equation, using a visual comparison process. The reasoning process generally starts at the highest relevant level of the part–whole hierarchy, viewing the entire left- and right-hand subexpressions as single chunks. If they do not match, then the reasoning process can descend the visual hierarchy through adjusting the spatial grouping of the perceptual system and reinspecting the equation (or a mental image of the equation). In this kind of model, the results demonstrated here follow naturally. Because perceptual grouping is the default method of order of operations evaluation, the presence of nonstandard grouping principles interferes with the effective mathematical procedure. We call an account like this an *interactive account*, because it assumes an interaction between higher level cognitive processes and an active visual apparatus.

Dehaene (Dehaene, 1997; Dehaene, Bossini, & Giraux, 1993; Dehaene, Molko, Cohen, & Wilson, 2004) has proposed a "triple-code" theory in which simple arithmetic facts are stored arithmetically, complex multidigit problems are represented visually, and numbers are compared through an amodal magnitude estimation process. Triple-code theory is primarily intended as an account of basic small-number arithmetic, and as such does not specifically address algebraic tasks that do not involve numeric computation. The mathematical task used here is beyond the scope of the current theory. In addition, triple-code theory has only been discussed in terms of single-grapheme atomic representations. However, triple-code theory is very well suited to account for the kinds of effects seen here in that it presents formal reasoning as explicitly grounded in specific, modal processes that are applied to rich visual representations. Although the current theory does not address or include specific visual processes other than symbol recognition, a model like triple-code theory could readily be expanded to include these groupings as a key component in formal interpretation without violating the existing theory.

Two-stage and interactive accounts of mathematical reasoning are very different, but the two are not incompatible; rather, we see interactive processes as built upon and including more commonly considered processes. Models such as ACT-R could be adapted to account for multigrapheme visual units through the overt inclusion of a visual grouping component, and rules that deal with them, without radical change in the overall architecture. We feel that inclusion of such processes would help formal models to address the pragmatic, though formally incorrect, visual strategies displayed by reasoners on this task.

### *Implications*

If multigrapheme perceptual units are the basic representations of much mathematical reasoning, then modelers ignore them at the risk of increased complexity in their models. Computational models typically separate symbol perception and identification from reasoning processes (such as determining operation order), which occur on abstracted representations of problem form. Such models structurally rule out the kinds of interactions we observe. This does not matter much for modeling basic single-operator arithmetic, perhaps, but it is important wherever formatting conventions are important and present. Models that ignore physical format risk modeling as conceptual interpretative processes that are really perceptual, and thus constructing processes over atomic graphemes that are not the basic conceptual units. Perceptual processes are strongly constrained; these constraints (if used) can provide modelers with valuable information about what symbols and/or systems can be easily learned, and how those systems are likely to be approached (Endress, Scholl, & Mehler, 2005).

These results also have practical implications for experimental work in algebra. Koedinger and MacLaren (2002), for instance, argued that symbolic expressions are more difficult than word equations for novice and intermediate algebraic reasoners. They suggest that many of these difficulties result from problems in interpreting algebraic equations; these difficulties lead to, among other errors, many failures to observe the correct order of operations. This is an important topic, and their experiments are well designed and compelling. Unfortunately, the symbolic problems presented to participants were uniformly spaced; our evidence suggests that many of the order of operations errors may have been induced by unhelpfully constructed symbolic equations, not from difficulties with symbolic expressions per se. Experimental claims about how symbolic information is acquired and used must in general carefully control, we suggest, the physical format of the presented equations. Researchers hoping to form general educational conclusions about higher level formal reasoning must pay close attention to the physical format of their stimuli, as well as to the formal content and presentation, when evaluating how reasoners engage with formal systems.

Understanding how people build complex relational understandings of mathematical forms is extremely important, both for building better mathematics instruction methods, and for improving mathematical notations. Although the effect sizes in Experiments 3 and 4 are not especially large, it is worth remembering that these effects existed despite feedback instructions to attend to the formal order of precedence rules. Also, in most practical situations, one does not simply make a single equality judgment, but instead makes such judgments while engaged in a more complicated task. Even in cases in which participants successfully made correct judgments, they may have used limited cognitive resources to do so; this could in turn impair performance on a larger task, of which building a syntactic representation is just a small part of the equation.

## 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. (2005). Human symbol manipulation within an integrated cognitive architecture. *Cognitive Science, 29,* 313–341.

Bassok, M. (1997). Two types of reliance on correlations between content and structure in reasoning about word problems. In L. English (Ed.), *Mathematical reasoning: Analogies, metaphors, and images* (pp. 221–246). Hillsdale, NJ: Erlbaum.

Bassok, M. (2001). Semantic alignments in mathematical word problems. In D. Gentner, K. J. Holyoak, & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 401–433). Cambridge, MA: MIT Press.

Bassok, M., Chase, V. M., & Martin, S. A. (1998). Adding apples and oranges: Alignment of semantic and formal knowledge. *Cognitive Psychology, 35,* 99–134.

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. *Quarterly Journal of Experimental Psychology 54A,* 1005–1029.

Campbell, J. I. D. (1994). Architectures for numerical cognition. *Cognition, 53*(1), 1–44.

Campbell, J. I. D. (1999). The surface form x problem-size interaction in cognitive arithmetic: Evidence against an encoding locus. *Cognition, 70*(2), B25–B33.

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

Dehaene, S. (1997). *The number sense*. New York: Oxford University Press.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122,* 371–396.

Dehaene, S., Molko, N., Cohen, L., & Wilson, A. (2004). Arithmetic and the brain. *Current Opinion in Neurobiology, 1,* 218–224.

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,* 406–419.

Fodor, J. A. (1975). *The language of thought*. New York: Crowell.

Gentner, D. (2003). Why we're so smart. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 195–235). Cambridge, MA: MIT Press.

Hadamard, J. (1949). *The psychology of invention in the mathematical field*. Princeton, NJ: Princeton University Press.

Harnad, S. (1990). The symbol grounding problem. *Physica D, 42,* 335–346.

Haugeland, J. (1985). *Artificial intelligence: The very idea*. Cambridge MA: MIT Press.
Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.
Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.
Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Sciences, 5,* 434–442.
Keil, F. C. (1989). *Concepts, kinds and development*. Cambridge, MA: Bradford Books/MIT Press.
Koedinger, K. R., & MacLaren, B. A. (2002). *Developing a pedagogical domain theory of early algebra problem solving* (CMU-HCII Tech. Report 02–100). Pittsburgh, PA: Carnegie Mellon University, School of Computer Science.
Koffka, K. (1935). *Principles of gestalt psychology*. New York: Harcourt Brace Jovanovich.
Lakoff, G., & Nũnez, R. E. (2000). *Where mathematics come from: How the embodied mind brings mathematics into being*. New York: Basic Books.
Landy, D., & Goldstone, R. L. (in press). Formal notations are diagrams: Evidence from a production task. *Memory & Cognition*.
Markman, A. B., & Dietrich, E. (2000a). Extending the classical view of representation. *Trends in Cognitive Sciences, 4,* 470–475.
Markman, A. B., & Dietrich, E. (2000b). In defence of representation. *Cognitive Psychology, 40,* 138–171.
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.
McNeil, N. M., & Alibali, M. W. (2005). Why won't you change your mind? Knowledge of operational patterns hinders learning and performance on equations. *Child Development, 76,* 883.
Morley, N. J., Evans, J. S. B. T., & Handley, S. J. (2004). Belief bias and figural bias in syllogistic reasoning. *Quarterly Journal of Experimental Psychology, 57A,* 666–692.
Palmer, S. E. (1992). Common Region: A new principle of perceptual grouping. *Cognitive Psychology, 24,* 436–447.
Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 324–354.
Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 629–639.
Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119,* 3–22.
Smith, E. E., & Sloman, S. A. (1994). Similarity- versus rule-based categorization. *Memory & Cognition, 22,* 377–386.
Stenning, K. (2002). *Seeing reason: Image and language in learning to think.* Oxford, England: Oxford University Press.
Stylianou, D. A. (2002). On the interaction of visualization and analysis: The negotiation of a visual representation in expert problem solving. *Journal of Mathematical Behavior, 21,* 303–317.
Tufte, E. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.
Tukey, J. W. (1972). Some graphic and semigraphic displays. In T. A. Bancroft (Ed.), *Statistical papers in honor of George W. Snedecor* (pp. 293–316). Ames: Iowa State University Press.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.
Wisniewski, E. J., & Bassok, M. (1999). Stimulus compatibility with comparison and integration. *Cognitive Psychology, 39,* 208–238.
Zazkis, R., Dubinsky, E., & Dautermann, J. (1996). Coordinating visual and analytic strategies: A study of the students' understanding of the group d4. *Journal for Research in Mathematics Education, 27*(4), 435–456.
Zhang, J., & Wang, H. (2005). The effect of external representations on numeric tasks. *Quarterly Journal of Experimental Psychology, 58A,* 817–838.

Received March 27, 2006
Revision received March 8, 2007
Accepted March 12, 2007 ■