Formal notations are diagrams: Evidence from a production task

DAVID LANDY AND ROBERT L. GOLDSTONE
Indiana University, Bloomington, Indiana

Although a general sense of the magnitude, quantity, or numerosity of objects is common in both untrained people and animals, the abilities to deal exactly with large quantities and to reason precisely in complex but well-specified situations—to behave formally, that is—are skills unique to people trained in symbolic notations. These symbolic notations typically employ complex, hierarchically embedded structures, which all extant analyses assume are constructed by concatenative, rule-based processes. The primary goal of this article is to establish, using behavioral measures on naturalistic tasks, that some of the same cognitive resources involved in representing spatial relations and proximities are also involved in representing symbolic notations—in short, that formal notations are a kind of diagram. We examined self-generated productions in the domains of handwritten arithmetic expressions and typewritten statements in a formal logic. In both tasks, we found substantial evidence for spatial representational schemes even in these highly symbolic domains.

It is clear that mathematical equations written in modern notation are, in general, visual forms and that they share some properties with diagrammatic or imagistic displays. Equations and mathematical expressions are often set off from the main text, use nonstandard characters and shapes, and deviate substantially from linear symbol placement. Furthermore, evidence indicates that at least some mathematical processing is sensitive to the particular visual form of its presentation notation (Cambell, 1999; McNeil & Alibali, 2004, 2005). Despite these facts, notational mathematical representation is typically considered sentential and is placed in opposition to diagrammatic representations in fields as diverse as education (Stylianou, 2002; Zazkis, 1996), philosophy of science (Galison, 1997; Perini, 2006), computer science (Iverson, 1980), and cognitive modeling and problem solving (Anderson, 2005; Stenning, 2002).

The standard conception of mathematical notation is best understood via Palmer's (1978) classic distinction between intrinsic and extrinsic representational schemes. A representation is intrinsic "whenever a representing relation has the same inherent constraints as its represented relation" (p. 271). For example, Line A's being shorter than Line B can be intrinsically represented by the representational element that corresponds to A's being shorter, taller, brighter, or larger than the element representing B—in other words, by any relation that is inherently asymmetric and transitive. Representations are extrinsic when their inherent structure is arbitrary. They model the represented world by explicitly building the structure that is needed to conform to the world. Palmer argued that analog representations are intrinsic, in that

correspondences and inferences between represented and representing worlds come for free because of their shared intrinsic structure. Propositional representations, including language, logic, and mathematics, are extrinsic and hence come to represent objects by explicitly establishing relations with whatever structure is needed. The only intrinsic relation necessary to propositions is the left–right concatenation of basic symbols. Although representations in mathematics and logic are traditionally understood as extrinsic, it is possible that they nonetheless possess intrinsic and analog properties, and it is this possibility that we empirically pursue here. In a separate study (Landy, Havas, Glenberg, & Goldstone, 2007), we consider the case for language.

Stenning (2002) tried to characterize the apparent distinction between diagrams on the one hand and formal equations and language on the other while also recognizing that both are frequently visual and schematic/abstract representational formats. Stenning proposed that diagrams represent relational structures directly, whereas notations—formal or otherwise—have structural information mediated via rules governing individual elements. That is, assuming that the represented domain consists of relation r governing objects in a set $\{a\}$ in a directly represented (diagrammatic) representation schema, there is a metric property of the representation R (such as spatial proximity) such that R and r act in direct proportion. In an indirectly represented language, r does not correspond to any metric feature of the representation. Instead, $r(a1, a2)$ is expressed via a set of rules governing concatenative strings, in which the only relevant property of a display is the order in which terms appear.

D. Landy, dlandy@indiana.edu

Specifically, we propose that formal notations are diagrammatic as well as sentential and that the property conventionally described as syntactic structure is cognitively mediated, in part, by spatial information. Elements of expressions are bound together through perceptual grouping, often induced by simple spatial proximity. Thus, our claim is that mathematical formalizations of syntax are not themselves the direct cognitive mechanisms typically employed in processing that syntax. The former really are concatenative, but we propose that people use space and spatial relationships in representational schemas to facilitate the processing of syntax. We are not claiming here that the execution of each individual step in a proof or computation is inherently spatial or processed exclusively using sensorimotor mechanisms. We do suggest that spatial reasoning with regard to the physical layout of notational forms is common in reasoning with formal languages and that spacing practices play a significant role in human reasoning using notations.

We have argued previously that a broadly similar interference of metric (non-order-related) spatial properties on syntactic judgments provides evidence that spatial processes and representations implement syntax in typical human judgments (Landy & Goldstone, 2007; see also Kirshner, 1989; Kirshner & Awtry, 2004). To study the influence of perceptual grouping on mathematical reasoning, we tasked undergraduate participants with judging whether an algebraic equality was necessarily true. The equalities were designed to test the students' abilities to apply the order of precedence of operations rules (e.g., the rule that multiplication precedes addition). Although our participants knew these rules, we were interested in whether perceptual and form-based groupings would be able to override their general knowledge of the order of precedence rules. We tested this by having grouping factors either consistent or inconsistent with order of precedence. For example, a participant might be asked whether $n*w + y*b$ was necessarily equal to $y*b + n*w$. In this example, the physical spacing around the operators was consistent with the order of operators, with more space around the plus symbols than the multiplication symbols. On other trials, the spacings were inconsistent. For instance, participants might be asked to judge whether $m+p * e+g$ was equal to $e+g * m+p$. (It was not.) When physical spacing was inconsistent with order of precedence rules, six times as many errors were made compared with when the spacing was consistent (neutral spacing trials were intermediate in accuracy). Participants continued to show large influences of spatial grouping on equation verification even though they received trial-by-trial feedback. This suggests that sensitivity to grouping is automatic or at least resistant to strategic, feedback-dependent control processes. Even for algebra, one of the clearest cases of widespread symbolic reasoning, spacing exerts a strong and lasting influence.

The fact that people are sensitive to small changes in the physical spacing of formal expressions suggests that symbol systems themselves are grounded, in part, in spatial and visual structures. If so, then people might also respond spatially to the syntax of internally represented expressions, leading symbolic productions to reflect syntactic structure—that is, the more tightly two mental terms are bound syntactically, the closer together people will tend to place them physically. A relationship between spatial and syntactic proximity would at best be unexplained if space is not part of how we represent formal syntax. Models of formal reasoning do not generally relate spatial properties and syntax (indeed, most definitions of formal reasoning overtly exclude visuospatial relations) and do not predict any correspondence between physical space and syntax (see in particular Anderson, 2005; Johnson-Laird, 1983). Nevertheless, if, as we suggest, understanding formal symbol structures typically involves spatial resources, then symbolic productions might be expected to reflect syntactic structure: The less tightly two adjacent terms are bound syntactically, the farther apart they should be placed physically. Experiment 1 directly examines our proposal by measuring physical interoperand spacing in handwritten equations constructed by participants from presented word equations.

EXPERIMENT 1

In this experiment, participants were asked to write out simple equations by hand. If, as we propose, formal notations automatically encode spatial relations corresponding to structural relations, then spacing in handwritten equations should reflect the formal structure of the equation. In particular, spacing around equality signs should be very large, since such signs denote, in all cases, the broadest partition of the sentence. Within the two-operator side of each equation, spacing should depend on the structure of the expression. In mixed expressions, the middle term is syntactically bound to the higher order multiplication sign, and so the spacing around that sign should be compressed (or that around the lower order addition sign expanded) relative to how it is spaced in unmixed expressions. Thus, we expect operator spacing to depend on the interaction between operator and context.

Although traditional sentential accounts of notational reasoning provide no reason to expect operator spacing to ever be nonuniform, experience with typeset or handwritten equations might drive some kinds of spacing regularities. Typeset equations are generally not of a fixed width, and multiplication symbols are generally narrower than addition signs. Experience with typeset equations could lead equation writers to generally space multiplicands more narrowly than addends. However, no prominent equation typesetter adjusts the spacing of terms on the basis of syntax (and if one did, we would regard this as evidence favoring our view). Therefore, although either the spatial encoding or the amodal perspective might predict a main effect of operator, neither the width of the operators themselves nor experience with typeset equations could lead to the predicted interaction. Thus, the principle theoretical measure is the interaction between operator sign and mixed versus unmixed structure.

Figure 1: Sample stimuli from Experiment 1. The first and second stimuli are mixed examples; the third and fourth are unmixed. Each set of operands appeared in all four operator contexts. For this experiment, the distance between operands on the single-operator side was ignored.

Method

Participants. Twenty four Indiana University undergraduates participated in the experiment as partial fulfillment of a course requirement. The experiment lasted about 25 min.

Procedure. Word equations were presented one at a time on a computer screen; the participants wrote out corresponding equations using standard mathematical symbols. The participants were instructed to use only standard Arabic numerals and formal operator symbols ($+$, $\times$, and $=$) and were explicitly asked not to use any parentheses. The participants were not asked to solve or evaluate the correctness of any equation, nor were they reminded of the correct order of operations.

For each participant, 10 sets comprising three numbers between 2 and 9 were randomly generated (because 1 is much narrower than other numbers, it was excluded from this experiment).

Syntactic binding between terms was systematically manipulated by altering the operator of equations. Each set of three numbers appeared in four different equations, one each with the operator structures plus–plus, plus–times, times–plus, and times–times, making 40 translations in all. One side of the equation contained the three numbers with two operators, and the other side of the equation contained the same expression, but with the first operation completed. Thus, if the set of numbers was {2, 4, 9}, the equations, in symbolic form, would be $6 + 9 = 2 + 4 + 9$; $2 + 36 = 2 + 4 \times 9$; $8 + 9 = 2 \times 4 + 9$; and $8 \times 9 = 2 \times 4 \times 9$. The middle two equations are labeled *mixed operator conditions*, whereas the first and last are termed *unmixed operator conditions* (since there is no particular hierarchical structure on either side of the equation). In 5 of the 10 equation sets, the expression containing three numbers and two operators appeared on the right side of the equation (as in the examples above), and in the other 5, it appeared on the left. This procedure eliminated any interference of particular number choices, since each production was compared with productions that were identical except for operator context.

Each participant received a different, randomly generated stimulus set. Word equations were presented one at a time on a computer screen and remained on the screen while participants wrote the corresponding symbolic equation in a printed box (1.1-cm high $\times$ 10.4-cm wide; see Figure 1) on a piece of paper. Word equations employed number words along with the words *times*, *plus*, and *equals*. For instance, if the word equation probe was "six plus five times four equals two plus nine times three," participants would respond by writing "$6 + 5 \times 4 = 2 + 9 \times 3$." Each participant viewed and responded to 40 equations in total. The paper responses were scanned, and responses that were left blank, contained parentheses or other extraneous marks, or contained crossed-out values or

other errors were dropped from the analysis. Then, the interoperand spacing between the symbols in the equations was measured using Adobe Photoshop.

We use *operand* throughout this article to refer to the two numerals on either side on an operator, regardless of operator precedence. Thus, in the expression $1 + 2 \times 3$, 1 and 2 are, for our purposes, the operands of the addition (the addends), even though normal mathematical usage would make 1 and 2×3 the operands. In addition, we will occasionally use the neologism *equalands* to refer to the numerals on either side of an equals sign. Although the equals sign is not usually thought of as an operator, it marks the highest level division of an equation, and so for our purposes serves a syntactic role akin to a very low order operation.

Results

For each participant, spacing was averaged across the stimuli in each condition. The mean values across participants for each context are shown in Table 1. The relationship between multiplications and additions was analyzed using a 2×2 within-participants ANOVA with distance as a dependent measure and operator and expression structure as independent variables. As predicted by the typesetting hypothesis, the ANOVA revealed a main effect of operator type [$F(1,23) = 7.89$, $MS_e = 3.35$, $p < .001$]. The interaction between operator type and expression structure was also significant [$F(1,23) = 4.73$, $MS_e = 1.28$, $p < .05$]. The compression of multiplicands relative to addends was greater in the mixed- than in the single-operator condition.

The relationship between equaland spacing and the spacing of other operators was explored using two independent t tests. Both contrasts were quite significant: Equalands were spaced more widely than multiplicands [$t(23) = 10.4$, $p < .0001$] and addends [$t(23) = 9.5$, $p < .0001$]. (A full 2×3 ANOVA including equalands verified the predicted effects but did not suggest any other significant interactions.)

Discussion

The most immediate conclusion that can be drawn from Experiment 1 is that syntax in arithmetic equations is processed automatically. The simple transcription task participants performed in this experiment required no consideration of syntax at all, but the results show a modulation of productions in response to syntactic structure. This result is somewhat outside our main focus but is interesting in its own right.

More interesting for our purposes is that the participants were not swayed arbitrarily by syntax—they constructed spatial properties that matched their own perceptual biases. Terms were spaced more narrowly when they were

Table 1
Mean Spacing (in Millimeters) by Measurement Condition in Experiment 1

	Unmixed Operators		Mixed Operators		Overall	
Operator	M	SE	M	SE	M	SE
Addition	9.52	±0.48	9.77	±0.48	9.65	±0.48
Multiplication	9.38	±0.50	9.16	±0.53	9.27	±0.51
Equality	12.38	±0.59	12.13	±0.57	12.25	±0.58

grouped more closely. Historical interactions with typeset equations do not predict these effects, nor do traditional symbolic accounts of mathematical competence. Thus, of the hypotheses considered earlier, this result is compatible only with the suggestion that people systematically vary spacing according to the particular syntactic structure of the current equation. Because this behavior presumably generalizes to the population at large, including the teachers and parents of our particular participants, historical interactions with other handwritten equations could account for the results—participants in our task might have been sensitive to syntax because their teachers and parents were, and therefore, sensitivity to the spacing of syntax formed part of the participants' training. That is, our participants may have received training with mathematical expressions in which the spacing was more often consistent with the syntactic structure than inconsistent with it. However, this explanation does not provide any additional insight into why this spacing convention has been adopted in the first place. For that, the most parsimonious account for the environmental regularity is, once again, that spatial processes are involved in the representation of mathematical syntax in the normal course of algebraic reasoning.

There is a plausible alternative to the spatial information hypothesis: It might be that syntax processing mediates (somehow) access to the lexicographic forms for numbers and symbols. If syntactically bound items are chunked in memory, for instance, then accessing terms within a chunk may be quicker than accessing terms across chunks (Cheng & Rojas-Anaya, 2006). If so, and if horizontal pen movement between characters correlates with access time (if, for instance, the pen is moved at some more or less fixed velocity while the lexical form is being accessed), then a memory delay could produce increased spacing. Experiment 2 addressed this possibility by exploring spacing behavior on a typed input task.

EXPERIMENT 2

A limitation of Experiment 1 was that the formal system used was a small (though important) one: prealgebra using equality, addition, and multiplication. Although this system is convenient in that it is widely known and studied, such a small system, with several idiosyncratic features, makes it difficult to generalize the results. Experiment 2 broadened the scope of our examinations by exploring a very different notational system: formal propositional (quantified and unquantified propositional) logic. In addition, Experiment 2 used a very different experimental format and a very different response generation system for the participants. Instead of asking participants to write unused and useless pseudo-equations in a laboratory setting, Experiment 2 involved a corpus analysis of expressions generated by participants interacting with a Web-based teaching tool designed and maintained by Colin Allen and Chris Menzel and based on an accompanying textbook (Allen & Hand, 2001).

Method

Materials. The current analysis is based on *Logic Daemon and Quizmaster* (Allen & Menzel, 2006, http://logic.tamu.edu/), an in-

teractive Web site designed for use with the *Logic Primer* textbook by Allen and Hand (2001). Students can use this Web site to tackle exercises found in the textbook as well as additional problems of the same type. For our analysis, we chose to focus on the translation exercises that require students to render sentences of English into the formal system described in chapters 1 (propositional) and 3 (predicate logic) of *Logic Primer*. In these exercises, students are presented with up to five English sentences; under each sentence is a standard Web form single-line text input field. Students freely type a response for one or more of the sentences and click a submit button. For instance, Problem 9 from problem section 1.3 (in chapter 1) states that "If Mary dances although John is not happy, Bill will dance." Instructions state that "Q," "S," and "R" are to be used to denote the atomic sentences "Mary dances," "John is happy," and "Bill dances," respectively. Participants then enter a formal sentence corresponding to this sentiment. [One correct response would be "((Q & ~S) -> R)."]

Each string of characters submitted in this way is checked first to see whether it represents a well-formed formula (WFF) according to the specifications of the formal system. Although the textbook uses non-ASCII characters, the character strings are mapped to ASCII strings for keyboard input. Specifically, the single arrow is represented with "->" (dash, greater than), the double arrow with "<->" (less than, dash, greater than), the upside-down A of universal quantification with "@," and the backward E of existential quantification with "$." The WFF formation rules specify the use of parentheses around binary sentential connectives: "&" (and), "v" (or), "->" (if . . . then), and "<->" (if and only if). Some of these parentheses may also be omitted following a formal convention that is defined in chapter 1 of the text. The parenthesis-dropping conventions follow the specified order of operations: "&" and "v" precede "->," and "->" precedes "<->." Any string that passes the WFF test is next checked for correctness with respect to the particular translation problem attempted (i.e., it is checked for logical equivalence to a stored answer). Both the WFF check and the correctness check are indifferent to any white space introduced by the student, and when problems are returned to students with feedback, any introduced spaces have been removed.

It is worth noting that although interface, formal system, physical situation, and participant pool and motivation were different in Experiment 1, the task is quite similar: In both cases, participants were asked to take a natural language statement and translate it into a formal system.

Analysis. A total of 129,526 submissions to the translation verification interface (Exercises 1.3 and 3.2) submitted between May 5, 2005, and April 4, 2006, were collected and analyzed. These raw submissions were reduced in several ways. First, the initial interface prompts a user to submit several translations at once; as a result, many users submit the same response to a single problem many times, because they continue to revise other problem responses on the same page. In order to eliminate repeated entries, repetitions of literally identical responses submitted from a single IP address on a particular day were eliminated. Also, many responses violated rules of the formal system that were neither syntactic nor semantic—by, for example, using incorrect symbols ("=>" instead of "->," "and" instead of "&," or "V" instead of "v"), failing to translate by simply retyping the English text prompt, and so forth. Since our overriding interest was in the relationship between syntax and space, submissions with invalid operator symbols were eliminated. Finally, many submissions were simply junk, such as "asdf," "dfdfdfdf." These were also eliminated. After all of these reductions, 48,131 statements from 595 unique IP address–time stamp combinations remained.

The translation verification interface automatically determines how well formed an expression that follows the formal syntax is. If the expression is well formed, the interface then determines whether it forms a correct answer to that problem. Both of these pieces of information are provided to the submitter, who then has the option to revise the submission. The same verification system used to provide submission feedback was used to categorize submissions for our analysis.

We distinguished three physical spacing conditions: spacing consistent with the operator structure, spacing inconsistent with operator structure, and no spacing at all (unspaced). An expression was considered consistent when the space around every operator in the expression was appropriate: Spaces around conjunction, disjunction, conditional, and biconditional signs should have been even, and spaces should have been inserted only to the left of negation signs and quantifiers. If any spacing violated these constraints, then the expression was flagged as inconsistent.

Our predictions were as follows: First, we predicted that since representations of space play a role in the way that reasoners process syntax, participants using the site would at least occasionally insert spaces. Although random insertion of spaces would be far more likely to produce inconsistent than consistent spacings, we predicted that spacing would primarily be consistent and that only consistent spacing would improve performance.

Experiment 1 indicated that spacing is modulated in the presence of hierarchical syntax. On this basis, we predicted that participants would be more likely to produce spaces in responses containing more than one operator. Because more skilled reasoners are less likely to be dependent on perceptual support (Chi, Feltovich, & Glaser, 1981), we also predicted that more advanced participants would be less likely to space expressions at all. The problems studied came from two sections of the book, one on propositional logic and one on predicate logic. Our prediction was that spacing would be more common in responses to problems in the first, more elementary section. Finally, because the spatial representation theory implies that consistent spacing is indicative of deep syntactic processing, we predicted that accuracy would be higher on consistently spaced statements than on other statements, whenever structure mattered (i.e., whenever there were two or more operators).

In analyzing these data, we did not attempt to evaluate the statistical significance of our results for two reasons: First, the breakdown of submissions by unique IP address and date did not adequately divide submissions into independent samples. Since we had no way to determine unique individuals, and furthermore no way to determine the relationship between individuals, statistical tests based on the assumption of independent samples were inappropriate. Second, the large size of the sample guaranteed that standard statistical measures would indicate significance (all of the contrasts considered here are highly significant by standard measures), regardless of the underlying mechanisms. For these reasons, we report frequencies without invalid statistical measures.

Results

Table 2 presents the frequencies of submissions broken down by spacing, number of operators, and logic type. As expected, participants frequently spaced expressions: 10.8% of all expressions submitted contained some spacing. When expressions were spaced, moreover, most were consistently spaced: 82.6% of all spaced equations were spaced consistently with operator syntax. Consistently spaced submissions were also more likely than either inconsistently spaced or unspaced expressions to be correct: 53% of consistent equations were correct, compared with only 37% of inconsistently spaced and 50% of unspaced equations.

In order to test the structure sensitivity of consistent spacing, we divided the data set according to whether or not a problem required syntax resolution (that is, whether it had two or more connectives; see Table 2). Participants did indeed space more frequently on multioperator problems (9.2% of multioperator problems were consistently spaced vs. 6.8% of few-operator problems). Furthermore, accuracy was highest (53%) when expressions were consistently spaced and lowest (37%) when they were inconsistently spaced (50.1% of all unspaced expressions were correct). This difference was larger on multioperator expressions. On few-operator problems, 52% of consistently spaced, 42% of inconsistently spaced, and 52% of unspaced submissions were correct. However, the number of such problems was very small (only 26 inconsistently spaced few-operator problems were submitted in all).

We also tested the theory that more training would reduce the need for formally extraneous spacing. Translation problems appear in two sections of the textbook: chapter 1 (propositional logic) and chapter 3 (predicate logic). We divided the full data set into these two categories and measured spacing frequency across these two categories (see Table 2). Both consistent and inconsistent spacings were more frequent on propositional problems (13.1% and 2.4%, respectively) than on predicate logic problems (6.7% and 1.6%, respectively).

Discussion

Despite being formally unnecessary and informally discouraged, spaces were frequently inserted into typed sentences of formal logic. These spacings were nearly always consistent with the operations they abutted; submissions with consistent spacing were also slightly more likely to be correct than unspaced submissions. Together with Experiment 1, Experiment 2 establishes that people working in two very different domains systematically spaced formal systems that formally do not require differential spacing.

Because the participants in this experiment were typing on a keyboard, a chunking account that predicts differential spacing on handwritten equations as a result of differential chunking in memory cannot account for spacing

Table 2
Use of Spacing in Typed Formal Translations in Experiment 2

| | Logic Type | | | | | |
| | Propositional | | | First Order | | |
	Unspaced	Consistent	Inconsistent	Unspaced	Consistant	Inconsistent
Single Operation						
Correct	1,344	160	11	1,187	12	0
Incorrect	1,592	145	15	735	9	0
Multiple Operations						
Correct	6,175	1,058	159	12,899	10,441	65
Incorrect	5,182	844	213	13,817	10,113	39

Note—Entries represent the number of submissions in each category.

here. The most plausible theory that predicts systematic spacing in both of these domains is that syntax processing is bound up with spatial representation.

Other accounts of the results of Experiment 2 are possible, however. For instance, other formal systems, such as programming languages, are often taught with explicit instructions to space logical terms; participants might have been transferring this practice from programming experience. This possibility cannot be definitively eliminated (although why spacing is common in programming is still mysterious), but the fact that more experienced reasoners spaced less, not more, seems incompatible with the idea that spacing appropriately is an acquired skill.

GENERAL DISCUSSION

In both typed logic and handwritten arithmetic translation tasks, participants created formally irrelevant spatial relationships in stimuli. In both cases, these relationships aligned with the syntactic structure of the formal statement being expressed. The kinds of regularities produced in Experiment 1 have been shown to benefit correct syntactic interpretation (Landy & Goldstone, 2007), and those of Experiment 2 seem at least to be correlated with more accurate submissions. People seem to spontaneously create alignments of space and syntax that, when perceived, help them reason formally.

Although behavioral evidence cannot conclusively establish that syntactic structure in mathematics is generally represented using spatial representational systems, it does establish a processing connection between space and syntax that is difficult to otherwise explain. Regardless of the internal structure of syntactic representations, the evident relationship between metric spatial properties and syntactic behavior indicates that the latter is, in Stenning's (2002) sense, a directly interpreted property of mathematical equations. Mathematical and logical forms often literally function as diagrams, serving to ground the abstract relationships they express in more immediately available sensorimotor relationships. In the normal course of events, spatial grounding of syntax guarantees spatial alignment with the judgments that purely concatenative rules demand.

Our claim is that external representations of syntactic systems carry formal information in their spatial properties. This is the converse of other arguments that hold that mathematical reasoning involves visuospatial processing rather than formal operations. For example, Hadamard (1949) complained that the true heart of a mathematical proof—the intuitive, often visual conceptualization—is ignored in the formal description of the proof steps themselves. The mismatch between how mathematics is formally presented and how it is concretely conceived by practitioners has been often noted (Lakoff & Nuñez, 2000). Our claim complements these analyses because both they and our approach seek to integrate perceptual processing and mathematical reasoning. However, our claim is that even when we consider symbolic transformations, visuospatial processes are still pertinent and strongly influential. Thus, unlike some proponents of embodied cognition (Wilensky

& Reisman, 2006), we do not advocate drawing a sharp division between perceptual and formal routes to mathematical understanding. We believe that formal routes to mathematical understanding are perceptual.

Space has been proposed as a mediating metaphor for metric properties such as time (Boroditsky, 2000; Casasanto & Boroditsky, 2003), numerosity (Dehaene, Bossini, & Giraux, 1993; Fischer, 2001, 2003; Lakoff & Nuñez, 2000), pitch (Casasanto, Philips, & Boroditsky, 2003), and many others (Casasanto & Lozano, 2006). Behavioral evidence typically taken as evidence for the spatial processing of temporal reasoning has included priming effects of spatial experience on temporal interpretations of spatial/temporal language (Boroditsky, 2000). Another traditional source of evidence for the role of spatial representations is the spontaneous production of systematic spatial behaviors that are not task appropriate. For instance, Casasanto and Lozano (2006) observed participants as they told preconstructed stories about events that contained nonspatial concepts for which spatial metaphors were common—for example, the concept of a stock's increase in value being expressed as "going up." Even when the language that the participants used did not contain a spatial metaphor, they often used spatial gestures that were compatible with the typical metaphor, indicating that the spatial metaphor was mediating their explanations. One possible explanation of our results is that representations of syntax are typically derived metaphorically from representations of space, just as those of time or numerosity seem to be. We are not certain that the data support such a strong conclusion. Our primary claim is that the rule systems that govern the interpretation of formal systems carry functional spatial information—in other words, they are diagrammatic. Metaphorical spatial derivations are, however, one plausible potential account of spatial representations of syntax.

That spacing facilitates syntax is important for our understanding of mathematics and mathematical learning, but it is also important for education and cognitive psychology more generally. For education, our results suggest that increased sensitivity to the physical features of how mathematics is presented to students and how the students in turn present mathematics to their teachers may be beneficial. Physical properties such as spacing may be used to give students a perceptual scaffold for the rules underlying algebra. Further research is necessary to know whether these scaffolds, when removed, help students to continue to obey the appropriate mathematical rules or if they act as crutches that thwart rule development. Reciprocally, by examining students' spacings of their own productions, we may be able to diagnose their misunderstandings. In the same way that manual gestures are sensitive indicators of inchoate explicit mathematical understandings (Alibali & Goldin-Meadow, 1993; Goldin-Meadow, Wein, & Chang, 1992), production spacing may indicate the beginnings, or lack thereof, of knowledge for order of precedence.

One possible source of the influence of syntax on space is that mental representations of formal syntax are internally spatial. Another possibility is that internal representations really are amodal and that the differential spatial

proximities between math or logic elements arise only during production. By this account, people have an amodal, extrinsic represented form that resembles standard symbolic conceptions. This representation incorporates syntax explicitly by formal structure rather than spacing. However, when participants produce physical forms in the production tasks, they are influenced by the syntax in their own spacings. This alternative remains a possible account of our results, and further explorations will be required to fix the source of the spatial–syntactic interaction, but several considerations make the production-based account unattractive. First, our earlier results showed that perceptual tasks, not just production tasks, were influenced by spacing (Landy & Goldstone, 2007). It is possible to explain both sets of results as being the result of peripheral changes to perceptual input and motor output processes, but it is more parsimonious to have a unified account of both perception and action in which representations are intrinsically spatial. Second, a related consideration of parsimony leads us to prefer a single representation that includes syntactic and spacing information, rather than positing two representations in which these sources of information are separated. We now know that spacing influences, and is influenced by, syntactic form, and thus the burden of proof may be said to be on those who argue that there is a deep syntactic representation that is completely isolatable from visuospatial information. Third, the production-based account leaves unanswered the question of why syntax influences spatial productions, whereas in our account, space is part of the represented formalism because it promotes correct syntactical parsing.

Following this third point, one likely reason that our participants showed sensitivity to syntactic form in their physical spacing of equations is that this spacing information promoted their own proper syntactic interpretation of the equations. We asked participants to produce equations without performing any additional operations on them, which is atypical. Participants usually write equations out in order to solve them. Even when our participants did not have to solve equations, they were apparently using the same production habits that they typically used when reasoning mathematically. Over the course of history, mathematical formalisms have evolved so that they are cognitively helpful devices, and this evolution has entailed making apparently superficial, but practically crucial, form changes. For example, the convention introduced by Descartes (Cajori, 1927), in which letters near the beginning of the alphabet are used to denote constants and those near the end to denote variables, frees us from the burden of remembering which are which and allows us to use our memory for other aspects of a mathematical situation. Similarly, using variable subscripts rather than distinct letters has no mathematical function at all; the only role of subscripts is cognitive: They help a reader to understand and remember which variables are related.

Our claim is that this is a case of cognitive ontogeny recapitulating phylogeny. The physical properties of mathematical formalisms are adapted not only across generations but also within a single person. People presumably learn to incorporate syntactically consistent spacing into

their physical equations because doing so helps them obey the formally sanctioned rules of mathematics. Even when these physical adaptations are discouraged (Experiment 2), they still persist, because people either learn by trial and error or intuit, via a metacognitive understanding of their own math skill, that formalisms that incorporate spacing end up serving their needs better.

Most fundamentally, our results challenge conceptions of symbols as amodal and divorced from analog, spatial information. In this respect, we offer a reinterpretation of Newell and Simon's (1963, 1976) influential physical symbol system hypothesis, which held that physical symbol systems had the necessary and sufficient means for producing intelligent action. A symbol system includes both physical symbols, such as marks on paper or punches on a computer tape, and the explicit rules for manipulating these tokens. In action, all of the physical symbols used in Newell and Simon's studies were distantly related to their worldly referents and were digital and discrete entities such as the strings "P ⊃ Q" and "GOAL 7 TRANSFORM L3 INTO LO." The arbitrary nature of these entities was by design, because Newell and Simon wanted the symbols they used to be capable of designating any expression whatsoever without any a priori prescriptions or limitations.

We concur with Newell and Simon's emphasis on physical symbols and believe in paying even more attention to symbols' physical attributes involving space, shape, and perceptual grouping. Accordingly, our revised physical symbol systems hypothesis is that symbols are not arbitrary, unconstrained tokens but rather are represented and processed using space and perceptually organized groups. This conception of physical symbols makes them far more constrained than those underlying Newell and Simon's General Problem Solver, but these constraints are not only limiters, but permitters as well. For specific problem solvers who are humans, it is good policy to design symbols that can be processed efficiently, given what we know about perceptual and cognitive mechanisms. From this perspective, it is hardly surprising if the symbols we write look a lot like symbols we are good at reading, and if the symbols we think with are a lot like symbols we are good at thinking.

AUTHOR NOTE

REFERENCES

ALLEN, C., & HAND, M. (2001). *Logic primer* (2nd ed.). Cambridge, MA: MIT Press.

ALLEN, C., & MENZEL, C. (2006) *Logic daemon and quizmaster*. Retrieved September 28, 2006, from logic.tamu.edu.

ALIBALI, M. W., & GOLDIN-MEADOW, S. (1993). Gesture–speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology*, **25**, 468-523.

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*, 1-28.

CAJORI, F. (1928). *A history of mathematical notations*. La Salle, IL: Open Court.

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

CASASANTO, D., & BORODITSKY, L. (2003). Do we think about time in terms of space? In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Meeting of the Cognitive Science Society* (pp. 216-221). Mahwah, NJ: Erlbaum.

CASASANTO, D., & LOZANO, S. (2006). Metaphor in the mind and hands. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 142-147). Mahwah, NJ: Erlbaum.

CASASANTO, D., PHILLIPS, W., & BORODITSKY, L. (2003). Do we think about music in terms of space: Metaphoric representation of musical pitch. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (p. 1323). Mahwah, NJ: Erlbaum.

CHENG, P., & ROJAS-ANAYA, H. (2006). A temporal signal reveals chunk structure in the writing of word phrases. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 160-165). Mahwah, NJ: Erlbaum.

CHI, M. T. H., FELTOVICH, P. J., & GLASER, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

DEHAENE, S., BOSSINI, S., & GIRAUX, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122*, 371-396.

FISCHER, M. H. (2001). Number processing induces spatial performance biases. *Neurology, 57*, 822-826.

FISCHER, M. H. (2003). Spatial representations in number processing—Evidence from a pointing task. *Visual Cognition, 10*, 493-508.

GALISON, P. (1997). *Image and logic*. Chicago: University of Chicago Press.

GOLDIN-MEADOW, S., WEIN, D., & CHANG, C. (1992). Assessing knowledge through gesture: Using children's hands to read their minds. *Cognition & Instruction, 9*, 201-219.

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

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

JOHNSON-LAIRD, P. (1983). *Mental models*. Cambridge, MA: Harvard University Press.

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

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

LAKOFF, G., & NUÑEZ, R. E. (2000). *Where mathematics come from: How the embodied mind brings mathematics into being*. New York: Basic Books.

LANDY, D., & GOLDSTONE, R. L. (2007). How abstract is symbolic thought? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 33*, 720-733.

LANDY, D., HAVAS, D., GLENBERG, A., & GOLDSTONE, R. L. (2007). *The role of perceptual groups in natural language interpretations*. Manuscript in preparation.

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.

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.

NEWELL, A., & SIMON, H. A. (1963). GPS: A program that simulates human thought. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought*. New York: McGraw-Hill.

NEWELL, A., & SIMON, H. A. (1976). Computer science as empirical enquiry: Symbols and search. *Communications of the ACM, 19*, 113-126.

PALMER, S. E. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 259-303). Hillsdale, NJ: Erlbaum.

PERINI, L. The truth in pictures. *Philosophy of Science, 72*, 262-285.

STENNING, K. (2002). *Seeing reason: Image and language in learning to think*. New York: 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.

WILENSKY, U., & REISMAN, K. (2006). Thinking like a wolf, a sheep or a firefly: Learning biology through constructing and testing computational theories—An embodied modeling approach. *Cognition & Instruction, 24*, 171-209.

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*, 435-456.