Journal of Experimental Psychology: Learning, Memory, and Cognition

Abstract Numeric Relations and the Visual Structure of Algebra

David Landy, David Brookes, and Ryan Smout

Online First Publication, May 12, 2014. http://dx.doi.org/10.1037/a0036823

Journal of Experimental Psychology:
Learning, Memory, and Cognition

0278-7393/14/$12.00 http://dx.doi.org/10.1037/a0036823

Abstract Numeric Relations and the Visual Structure of Algebra

David Landy
Indiana University and University of Richmond

David Brookes
Florida International University

Ryan Smout
University of Richmond

Formal algebras are among the most powerful and general mechanisms for expressing quantitative relational statements; yet, even university engineering students, who are relatively proficient with algebraic manipulation, struggle with and often fail to correctly deploy basic aspects of algebraic notation (Clement, 1982). In the cognitive tradition, it has often been assumed that skilled users of these formalisms treat situations in terms of semantic properties encoded in an abstract syntax that governs the use of notation without particular regard to the details of the physical structure of the equation itself (Anderson, 2005; Hegarty, Mayer, & Monk, 1995). We explore how the notational structure of verbal descriptions or algebraic equations (e.g., the spatial proximity of certain words or the visual alignment of numbers and symbols in an equation) plays a role in the process of interpreting or constructing symbolic equations. We propose in particular that construction processes involve an alignment of notational structures across representation systems, biasing reasoners toward the selection of formal notations that maintain the visuospatial structure of source representations. For example, in the statement "There are 5 elephants for every 3 rhinoceroses," the spatial proximity of *5* and *elephants* and *3* and *rhinoceroses* will bias reasoners to write the incorrect expression $5E = 3R$, because that expression maintains the spatial relationships encoded in the source representation. In 3 experiments, participants constructed equations with given structure, based on story problems with a variety of phrasings. We demonstrate how the notational alignment approach accounts naturally for a variety of previously reported phenomena in equation construction and successfully predicts error patterns that are not accounted for by prior explanations, such as the left to right transcription heuristic.

Keywords: mathematical cognition, symbolic reasoning, analogy

Mathematics is the study of abstract relations among abstract entities. Even something as simple as addition involves abstraction; the same operation applies whether we are adding grains of sand or galaxies. Among the wide variety of possible mathematical representations, algebraic equations are one of the most powerful and ubiquitous means of expressing quantitative relationships. Formal algebraic statements in which variables are used to represent relationships among unknown or indeterminate quantities express some of the most abstract assertions many people likely ever consider.

David Landy, Department of Psychological and Brain Sciences, Indiana University, and Department of Psychology, University of Richmond; David Brookes, Department of Physics, Florida International University; Ryan Smout, Department of Psychology, University of Richmond.

This research was funded by Department of Education, Institute of Education Sciences Grants R305H050116 and R305A110060, as well as a research grant from the University of Richmond. We thank Mark Ashcraft, Miriam Bassok, Tom Carr, John Hummel, Jose Mestre, Lydia Nichols, Brian Ross, Eric Taylor, Dylan Vavra, and the members of the Cave Lab for useful suggestions and ideas.

Correspondence concerning this article should be addressed to David Landy, Department of Psychological and Brain Sciences, 1101 East 10th Street, Indiana University, Bloomington, IN 47405. E-mail: dlandy@indiana.edu

Algebraic expressions may encode abstract relationships but are themselves physical forms. Although notations (any written inscription that encodes meaning) often have fewer salient surface features than many other symbol systems (Kaminski, Sloutsky, & Heckler, 2008), formal notations such as algebra, diagrams, and charts typically make extensive use of physical proximity and spatial relations. Recent work has repeatedly demonstrated that detailed physical properties and relations—and the perceptual systems that process them—play an important role in reasoning using visual displays in a variety of abstract contexts including puzzles (Patsenko & Altmann, 2010), biological taxa (Novick & Catley, 2007), weather diagrams (Hegarty, Canham, & Fabrikant, 2010), and algebra (Kirshner, 1989).

In the case of symbolic algebra, a tension between formal syntactic rules and visual patterns can arise. The formal syntax of algebra encodes relations among quantities by abstract rules specified in terms of ordinal relations among symbol tokens. The same mathematical expression can be rearranged in a number of ways without altering the formal content of the equation. On the other hand, a particular symbol system, such as mathematics, uses spatial relations in unique ways to convey additional meaning (Sherin, 2001). In the usual algebraic syntax, this tension creates visual structures that are sometimes aligned with (Landy & Goldstone, 2007b) and sometimes opposed to (Kirshner, 1989; Kirshner & Awtry, 2004; Marquis, 1988) underlying mathematical content. A rich notational calculus such as modern algebra replaces some conceptual relations with concrete organizational relations specified by principles of visualization and perceptual organization (Dantzig, 1930; Goldstone, Landy, & Son, 2010; Malle, 1993). For example, the relatively large space surrounding the equals sign in an arithmetic equation allows spatial grouping to serve as a proxy for the conceptual separation of an equation into left-hand and right-hand sides (Landy & Goldstone, 2007a). Our purpose in this paper is to explore the role of spatial relations such as proximity and ordering in the construction of notational models (e.g., expressions and equations).

Using the Structure of Notations to Construct Equations

How do reasoners use algebraic notation to solve practical problems? This paper focuses on one essential step: the creation of a notational expression—in this case, an equation—that models a situation. One possible account suggests that much of the time this equation-construction process involves selecting an appropriate template—say, a memorized equation pattern such as $F = ma$, $y = mx + b$, $ax^2 + bx + c = 0$, or *base + change* (Carlson-Radvansky & Logan, 1997; Fisher, Borchert, & Bassok, 2011; Sherin, 2001)—and applying structure mapping (Gentner, 1983; Hummel & Holyoak, 1997) to align that structure with other representations of the scene, such as written descriptions, pictures, or mental simulations (Bassok, Chase, & Martin, 1998). For example, asked to find the total weight of a case with 6 apples, each apple weighing 10 ounces, one might retrieve an *a times b* frame and fill in the mapping $a = 6$, $b = 10$ via alignment to yield 6^*10.

The selection and use of equations will interact with the selection and use of other kinds of diagrams and mental simulations. A person reads a description and, on the basis of this reading and past experience, constructs mental simulations that may potentially model the situation. These may include images, conventionalized structures such as number lines, or candidate expression fragments adapted or selected from those previously encountered (Fisher et al., 2011; Sherin, 2001). These simulations are made specific by a process of alignment: We suggest that structure mapping (Gentner, 1983) is used to identify entities in text or images with those in equations. Once multiple representations have been constructed and specified, it is possible to “check the answer” by verifying the validity of the analogy after performing analogous transformations across multiple representations (e.g., filling a specific number into both an expression and a sentence and making sure both yield identical results).

This alignment-based account generalizes the semantic alignments reported by Bassok and colleagues in arithmetic (e.g., Bassok, Wu, & Olseth, 1995). Bassok et al. (1998) reported substantial evidence for alignment between the semantic aspects of source descriptions and particular arithmetic operations: For example, people are much better at producing story problems that involve the division of flowers across vases than story problems that involve their sum. This is interpreted as occurring because the asymmetric mathematical structure of division aligns well with the asymmetric semantic relation active in typical vase–flower interactions.

Semantic alignment posits that the construction of a formal model involves an alignment across situational representations: aligning semantic relations involved in the story and mathematical relations between operations, based on attributes of those relations such as whether the relation is symmetric or asymmetric. What about the physical aspects of the notation itself? The physical objects that compose an equation carry their own sets of properties (sometimes called attributes or features) and relations. For instance, they must be written in some color and with some size. They may be written poorly or written well. The same holds for a story description: In the sentence “There were four vases for each white rose,” the fact that the entities are roses, rather than some other type of flower, is likely a superficial aspect of the situation (for the particular purpose of writing a relational equation); the fact that the sentence contains only one multisyllable word is a property of the notation. The notation also carries structural or relational information. Some of this information expresses ordinal relations or even the spatial layout of the sentence. In the sentence above, for instance, the word *vase* appears earlier in the sentence than the word *rose*. Other notational aspects involve the hierarchical syntactic analysis of the sentence, as, for example, when the word *white* modifies the word *rose*. Though one might think of these as superficial aspects because they appear in the analysis of the notation rather than the intended meaning, they are not superficial in the same way that the specific color of the rose is. Some of the properties and relations in the notation will be important; for instance, when parsing the sentence itself (others of course may not be). We will call aspects (relations and properties) of the notation *notational structure,* to distinguish superficial semantic from notational information. When a story is mapped into an equation, there are two relevant sets of notational information: that carried by the story text and that carried by the equation.

Structure mapping involves a search for high-quality mappings between situations or representations (Gentner, 1983). One-to-one mappings across representations are high quality when they align objects along multiple collections of properties and relations in a structurally consistent manner. We suggest that both aspects of signified situations (semantics) and the physical aspects of notations themselves are used in the alignment process. Consequently, the highest quality mappings will be achieved when these properties and relations all align to suggest one mapping. Previous explanations have not considered the potential role of notational properties and relations in the alignment process. This account predicts that mappings will be preferred when they place symbolic tokens into spatial relations analogous to the spatial relations of a source image or the linguistic or spatial relations of a source text. It should be easier for people to connect “John had two stamps and got three more” to the arithmetic expression $2 + 3$ than to the expression $3 + 2$. Both expressions express the same meaning, but because the first case also maintains the left-right ordering, the resulting mapping is of overall higher quality than that between the text and the second case. Beyond including simple left-right ordering, we will include physical proximity, similarity, and other grouping relations that have been shown to be important in the interpretation of language (Bever, Jandreau, Burwell, Kaplan, & Zaenan, 1990) and in mathematical syntax (Kirshner, 1989; Landy & Goldstone, 2007b). Although in simple examples it may be easy to invert the order, in more complex situations this analogical mapping may become quite difficult. Consider for example a sentence from our Experiment 2: *There are five rhinos for every six elephants*. We suggest that this sentence is readily aligned with an expression such as $R/5 = E/6$. The spatial organization of this expression maintains the bindings between *five* and *rhinos* and *six* and *elephants*. In the case of the correct expression in multiplicative form $R^*6 = E^*5$, the spatial organization of the equation is in conflict with the sentence. In our account, it becomes challenging for reasoners to find a consistent mapping from the sentence to the equation because of the structural differences between the verbal representation and the mathematical one. In general, semantic and notational factors may reinforce or contradict each other; each is predicted to impact problem difficulty.

Terms are considered grouped in stories to the degree that that they are closely linked in the hierarchical syntactic analysis of the sentence. Words are grouped in equations when they are bound by high-order terms or appear together in the same visual space. For instance, in the sentence *The number of ducks times five is six times the number of swans, of* and *ducks* are tightly bound syntactically, while *ducks* and *times* are less tightly bound but are still bound more tightly than are *five* and *six*. Although *five* and *six* are physically proximal, syntactically they lie in distinct arguments (and on opposite sides) of the main verb of the sentence. Previous work by Reed (1987) and Weaver and Kintsch (1992) indicated that grammatical and narrative structures of word problems are spontaneously aligned by students when comparing stories or solving sets of stories. In this paper, only organizational and syntactic notational relations will be considered, and we specifically focus on alignments across rather than within representation schemes. Other notational aspects (e.g., color, font, size, or letter similarity) may play an important role in notational alignment, but this paper focuses on structural proximity.

The Left-Right Transcription Heuristic

Some instances of alignment effects have been found in prior work on equation production. In the so-called *students and professors* problems, participants produce an equation modeling a sentence relating two entities such as "There are six students for every professor" (the actual entities in the story may or may not be students and professors; the Appendix presents materials used in the current experiment). Because the correct multiplicative expression, $6P = S$, inverts the structural relations in the sentence notation (i.e., the sentence puts *six* and *students* in the same phrase, while the equation connects *6* and *P*), this pair has low equation alignability, and, indeed, has been shown to be very difficult to produce correctly (Clement, 1982; Clement, Lochhead, & Monk, 1981; Hegarty, Mayer, & Monk, 1995; Martin & Bassok, 2005; Mestre & Lochhead, 1983). A response in which the relationship between two variables in a simple proportional or additive relationship are swapped (in this case, $6S = P$) is called a *reversal error*, or simply a *reversal*. Our account makes the general prediction that expressions that align aspects of notation—such as proximity, arrangement, and symbol similarity—across representation forms will be more often produced and more easily interpreted. Assuming that expression forms are selected on the basis of general familiarity as well as apparent appropriateness, problems will be solved more accurately when the notational structures of readily accessible expression structures are aligned with correct responses.

Although we present reversal errors as a type of alignment effect, these errors have usually not been interpreted as instances of deep structure mapping across notations. Instead, they have often been attributed to the simpler *left-right transcription heuristic*. In this heuristic, the expression is built up purely from keywords such as "less than" or "times" embedded in a story text, with the general constraint that the result be a syntactic expression (Clement, 1982; Clement et al., 1981; Hegarty et al., 1995; Hinsley, Hayes, & Simon, 1977; Paige & Simon, 1966), without any processing of the "meaning" of the story. For instance, consider a problem statement like "John had some pennies. He gave away three pennies, and after that he had as many as Mary." In this case, a person using a left-right transcription process would use only the keywords "John . . . gave away . . . three . . . as many as . . . Mary." Converting each keyword yields $J - 3 = M$, without any consideration of the meaning of any of the words.

The left-right transcription heuristic is usually described as a heuristic done in place of modeling when possible. When it is not possible, modeling must be engaged. Because this style of explanation invokes distinct processes separately responsible for reversal errors (transcription) and correct responses (modeling), we will call it the *dual route* account. In this account, the representations of models and heuristics are quite distinct: While modeling does not retain notational properties such as expression order, left-right translation works only in terms of them.

The dual route account serves as a baseline model against which to evaluate the argument for incorporating notation alignment into modeling processes, reflecting a default rather than a specific theoretical commitment of earlier work. Because our account situates notational and semantic alignment as core processes of notation construction, we predict it to have effects even when modeling is engaged. On this account, reversals may sometimes occur because of transcription but are often or usually the result of alignment processes embedded within modeling itself. Thus, it is worth examining cases where the process of transcription is blocked.

One occurs when pieces of the equation are not physically present in the explicit form of the text. For instance, a sentence like "write a quadratic equation with coefficients of 6, 2, and 9" does not fully reflect the appropriate equation, $6x^2 + 2x + 9 = 0$. Instead, that equation can only be reached through a process of semantic interpretation, retrieval from memory, and, indeed, some guesswork. Thus, the process of transcribing as usually defined is impossible in this case; one simply cannot solve the problem without engaging the meaning of the text. However, the concrete inscriptions align nicely, in that the coefficients are bound in a common phrase in the sentence and appear in the same order bound together on one side of the resulting equation (which is in a default format). The selection of a common template will favor regularly used forms, while the alignment process will tend to put coefficients in the same order and structure as in the phrase. Experiment 1 tests a case similar to this but using Newton's gravitational equation.

A second case in which strict transcription is blocked is when the equation involves vertical structure, such as division. Recently, Fisher et al. (2011) demonstrated that the reversal rate of typical students-and-professors-style problems sharply decreases when students are forced to use an equation structure that creates a high-quality alignment between text and equation. In their experiments, students were encouraged to construct division models such as

$$\frac{S}{6} = P$$

for sentences such as "There are six students for every professor." Fisher et al. interpreted this result assuming that the nonstandard vertical structure blocked simple transcription heuristics, forcing students to engage in (more successful) modeling.

It is worth noting that the reason for the decrease in transcription errors, though central here, is peripheral to the primary purpose of Fisher et al. (2011). Their major contribution lies in revealing how students in algebra rely on multiplication rather than division as a standard form for equations, even when they would be better off using division. We entirely agree with this important conclusion and replicate it in Experiments 2–3, below. We have a somewhat different interpretation of the correct behavior in division cases. While Fisher et al. assumed that transcription heuristics are used in multiplication, but not in division, our *notation alignment* account suggests that in both cases modeling will occur, and an important component of that modeling will be the alignment of notational structures between the story and the equation. Modeling will be more successful, in our account, when the preferred or selected concrete forms align with notational (i.e., textual) properties of stories. In line with this, note that while the correct multiplicative statement $6P = S$ does not align notational relationships among the tokens of the sentence and the expression in the example above, the division equation does align them. In line with our account, students encouraged to use the division expression made many fewer reversal errors than those encouraged to use multiplication and gave substantial evidence of modeling.

As another example, consider a problem like "There are five elephants for every three rhinoceroses. Write an equation relating the number of elephants to the number of rhinoceroses." The story notation puts *five* grammatically close to *elephants* and *three* grammatically close to *rhinoceroses*, which together with a default bias toward multiplication predicts an equation like $5E = 3R$ (i.e., a semantic reversal). If the problem is solved via a division equation, the dual route and notation alignment accounts make similar predictions. The process of direct transcription is impossible here, so modeling should be necessary. If transcribing is primarily responsible for errors, then errors will be reduced. If the alignment of notation is responsible for reversals, errors will also be reduced here but for a different reason: The division equations that put 5 and E into alignment and 3 and R in alignment include $\frac{E}{5} = \frac{R}{3}$ and $\frac{5}{E} = \frac{3}{R}$, which align "left side" with the first phrase of the story and "right side" with the second, or even $\frac{E}{R} = \frac{5}{3}$, which aligns "numerator" (or "top") with the first side and "denominator" with the second. Of importance, all these equations are correct; so aligning notation tends to produce correct responses.

In other cases, these two accounts make different predictions. Consider "There are nails on one side of a balanced scale, and screws on the other. There are 9 nails for every 4 screws. Using N for the weight of one nail and S as the weight of one screw, write an equation that expresses the relationship among the weights of the nails and the screws." On a balanced scale, weights are proportional inversely to counts, because the total weight on each side is the product of the number of items and the weight of a single item. Therefore, one can apply the proportionality schema to solve the problem (Inhelder & Piaget, 1958). In this example, the screws must be heavier than nails, because it takes fewer of them to balance the larger number of nails. At this point, the problem is one of proportionality, just as when variables are counts. Thus, a correct equation would be $9N = 4S$. Thus, transcription produces correct answers, as would the notation alignment process, when the equation is multiplicative. If the equation produced involves division, on the dual route account transcription processes are blocked, so again accuracy should be high. However, the notation alignment account predicts a different set of outcomes. Errors should indeed be rare in the multiplicative case, but they should be frequent in the case of a division equation. This is because the aligned division equations, such as $\frac{9}{N} = \frac{4}{S}$, reverse the correct relation. This case is directly probed in Experiments 2 and 3.

We do not suggest that that left-right transcription heuristics never occur, but our notational alignment account does not attribute the bulk of reversals to shallow errors. Instead, reversals such as those discussed above will be common whenever the syntax of English language statements and spatial structure of appropriate mathematical expressions are misaligned—whether or not a strict left-right transcription process is likely or possible. Reversals occur because the process of modeling itself involves alignments across notational schemes, and surface-level matches facilitate these alignments. The alignment process we suggest also predicts left-right biases in the cases where it is usually found, but we see such alignment as part of normative equation construction, not just as the result of a separate and shallow heuristic. We predict that ordering will affect outcomes even in cases where strict transcription processes are impossible to apply and modeling is necessary for equation construction.

To sum up: Traditional explanations based on transcription heuristics parcel the effects of notation and of meaning into two dual routes, with little or no overlap. The notational alignment account we suggest here integrates them: Semantic and notational alignment is an essential step in model-based reasoning. Because of this, the notational alignment account predicts that even when equations must be retrieved from memory through a meaning-driven process, or otherwise cannot be strictly transcribed from the words of the problem, word and equation will tend to align into a common structure.

In three experiments, we evaluate the role of the notational alignment account in several novel equation production contexts. Experiment 1 conceptually adapts Chatterjee, Southwood, and Basilico's (1999) demonstration of the tendency to align across language and pictures to equation production, exploring whether responses are aligned even when the form of the problem lacks elements of the equation, which must be retrieved through modeling. Experiments 2–3 systematically manipulate the alignment of structural notational relations across text and equations in students-and-professors-type problems, exploring cases for which the notational alignment account makes clear predictions about relative difficulty. In most of these cases, strict left-right transcription as defined above is impossible.

Experiment 1

Experiment 1 explores the existence of a spatial–linguistic alignment bias in the construction of equations from text, in a case in which the equation must be constructed from memory (i.e., the full equation cannot be extracted from the text). Our procedure is adapted from experiments by Chatterjee et al. (1999) and Maass and Russo (2003) demonstrating that the concrete characteristics of a source sentence influence, to some degree, the preferred arrangement of elements of a picture. For instance, when drawing pictures of sentences such as "the girl pulled the boy," speakers of left-right-oriented languages tended to draw the girl to the left of the boy.

We asked participants model a story by constructing a Newtonian gravitational equation:

$$F = G\frac{m_1 m_2}{r^2}.$$

The key question was whether participants placed the mass terms in the equation in the same order in which they were introduced in a problem statement. This provides a good analogy to experiments demonstrating concrete notation alignment in the language–picture context, because either order is correct. More important, the gravitational equation contains several terms (the constant G and the square of the distance between the objects) and operations that are not specified by the lexical items in the text. Because only the mass terms appear in the story (see the Appendix for the specific materials used), strict application of the left-right transcription heuristic in the absence of modeling would lead to an incoherent formal expression, such as $m_1 m_2$. Transcription is not predicted when it yields incoherent results, so the baseline dual route model predicts modeling (or, potentially, other heuristics) in this case.

Method

Participants. Participants were 32 undergraduates at the University of Illinois who had recently completed introductory physics and who participated in exchange for monetary compensation. Introductory physics students were chosen because the materials assumed familiarity with gravitation equations. Participants were recruited through fliers placed around the physics building and near classrooms. Participants completed several problems related to other experiments. The number of participants was selected to be appropriate for those other experiments.

Procedure. Participants completed a short test containing several elementary mechanics problems and other distractors. Some problems were parts of other experiments, which will not be discussed here. The target problem was the fifth problem in a set of 16. In this problem, participants were told a story about several asteroids (with masses *m1, m2,* and *m3*) and a single asteroid (mass *m*). The order in which the terms were introduced was manipulated: In one condition (the *single asteroid first* condition), the single asteroid was described first (it moved from asteroid to asteroid). In the other condition (*several asteroids first*), the several asteroids were introduced first and were described as moving past the single asteroid.

The other items in the study involved physics problem solving, in fairly typical situations that a student might see in a classroom. Some had previously been used as exam or test questions in other classes. The content of these problems was unrelated to that of the target problem; no problems other than Problem 5 involved celestial bodies or gravitation.

The participants were asked to construct the Newtonian gravitation equation for the pairings of the single asteroid with each of the several asteroids. Participants were not reminded of the equation form; they had to reconstruct it or retrieve it from memory on the basis of semantic modeling. Assuming an appropriate frame was reconstructed, the participant must decide which mass term to place on the left and which to place on the right in the numerator of the gravitation equation. Because simply transcribing in this case would lead to gibberish, the dual-route model does not make any strong prediction about ordering of the terms. If the alignment of notational structure is a core component of equation construction processes, though, participants should tend to place terms in the order in which they were introduced in the problem.

Results

Responses were analyzed except when the item was left blank, only one mass term was included, or it was impossible to identify the mass terms. Of the participants, 27 either responded correctly or made only minor errors (typically neglecting to square the distance in the denominator). Four participants in the single asteroid first condition and one participant in the several asteroids first condition either failed to respond or wrote extremely incorrect answers, which did not involve two mass terms. All participants whose responses could be analyzed placed the mass terms consistently across the three gravitation equations. Eighty-nine percent (12 in each condition) placed the terms in the equation in the order in which they were introduced in the narrative ($p < .001$ by Fisher's exact test).

Discussion

Our purpose in this experiment was to examine whether reasoners match order structures across equations and source narratives, particularly in cases in which modeling is necessary in order to generate an equation but the order does not matter (as is the case with the masses of the two asteroids in the gravitational equation). We found a powerful tendency toward ordinal alignment similar to that previously found in a picture-drawing task. Modeling is needed to get to the right equation, so notation-alignment pressures and modeling must coexist in solutions to the same problem.

A general pressure toward matching notation order in clear modeling contexts is a necessary precondition for the notation alignment account. The existence of such a pressure contradicts the baseline dual route account described above, because in that account left-right transcription is a separate process from modeling, and left-right transcription is impossible here. This result does not on its own indicate that alignment is an important factor in the production of equations. The flavor of the dual-route account is that modeling generally produces correct expressions, while reversal errors are usually caused by flawed heuristics not based on modeling. It might be that there is some weak pressure in favor of alignment that operates only when other factors are absent, but nevertheless be the case that most errors are caused by separate heuristics such as left-right alignment.

Gravity involves an interaction between two objects. Each asteroid exerts the same gravitational force on the other, if the situation is appropriately modeled. The instructions asked participants to write an equation that expressed the mutual gravitational force between them, but in each case, one asteroid set was described first, moving past the other. It is therefore possible that some participants inappropriately inferred a causal asymmetry from the order or the movement, and that some of the alignment seen here results indirectly from the known bias to put causes to the left of effects (Mochon & Sloman, 2004). Alternatively, it is quite plausible that these biases are themselves a result of notational alignment, together with the well-established tendency to consider actions in visual scenes as proceeding rightward, with causes on the left (Chatterjee et al., 1999; Kranjec, Lehet, Bromberger, & Chatterjee, 2010).

Experiments 2 and 3 explore predictions of notation alignment in traditional students-and-professors contexts. In any account of equation production, these problems lead to many reversals because of an unfortunate mismatch between the structure of common comparative expressions in algebra and the structure of common comparisons in English. There is no deep reason why algebra and English should be structured in this way, but they are. In a dual route account, this mismatch leads to errors because of inappropriate application of heuristics; thus, errors should be reduced by any intervention that reduces the rate of inappropriate heuristic use, such as using language phrasings or equation structures that do not admit easy transcription. On the notation alignment account, structure mapping is a core component of equation modeling; errors should be reduced when the text and the correct equation align and should be higher when they mismatch across contexts. Table 1 presents several examples of alternative ways to phrase comparative statements in English, and multiplication and division equations. Each sentence is accompanied by two models based on inverse operations in which variables stand in for the magnitude of the object sets referred to. In each case, one operation aligns concrete notational structures, while the other misaligns them. For instance, the relational statement "There are five rhinos for every six elephants" is correctly modeled by both the multiplicative statement $6R = 5E$ and the division statement $R/5 = E/6$. Neither maintains the left-right order of the sentence, but the division model maintains the proximity relationships of the text notation; thus, notational structure of the division model and the sentence are better aligned.

Experiment 2 provides a preliminary replication of Experiment 1 of Fisher et al. (2011), in which participants were required to produce multiplication or division equations. Experiment 2 extends prior work in two ways: First, Experiment 2 used a new manipulation to create a predicted interaction within division problems: Sometimes the variables represented the number of objects (as in traditional students and professors problems), but sometimes the variables represented the weight of a single item, and reasoners wrote an equation using variables to stand in for the weight of each object. Because mathematically in this situation numbers and weights relate inversely, this has the effect of inverting the concrete relations of the correct equation for a particular text, compared to the more typical form requested in relational equations. For example, if there were five rhinos for every six elephants on a balanced scale, $5R = 6E$ is a correct model of the weights involved (because five times the weight of a single rhino really would be the same as six times the weight of a single elephant). *Number* problems asked participants to solve the more typically explored equation in which variables stand for set sizes. Dual route accounts can naturally predict that modeling weights would be, say, more or less difficult than modeling item counts, which would lead to a main effect in variable type when modeling is engaged (e.g., in division problems). However, only the notation alignment account predicts a three-way interaction among variable, phrasing, and frame, and in particular a two-way interaction within just the division problems, such that particular phrasings make a problem difficult exactly when they misalign the surface structure and the correct response.

Experiment 2

Method

Participants. Because our goals in Experiment 2 did not rely on physics expertise, a more general population was used in Experiment 2 than in Experiment 1 (in this case, undergraduates taking introductory psychology). Sixteen undergraduates attending the University of Richmond received partial course credit for participation. On the basis of pilot testing, we believed that this population size would be sufficient to evaluate the predictions of the notation alignment account.

Design. We constructed several relational equation problems. Each described in a short paragraph (2–4 sentences) an equilibrium condition of two similar objects (e.g., screws and nails) on a balanced scale. The critical sentence numerically described this relationship using two relatively prime constants. For instance, the critical sentence of one problem asserted, "Measuring these ingredients on a balance, [Janet] notices that for every seven chocolate chips on one side of the balanced scale, there are five nuts on the other side."

Table 1
Sample Phrasings, With Multiplication and Division Models With Number Variables

Phrasing type	Examples	Multiplication model	Division model
Direct comparison	There are four screws for every nail.	$4N = S$	$\frac{S}{4} = N$
Direct comparison	There are five rhinos for every six elephants.	$5E = 6R$	$\frac{E}{6} = \frac{R}{5}$
Hypothetical comparison	If there were four nails for every nail there actually is, there would be as many screws as nails.	$4N = S$	$\frac{S}{4} = N$
Operation statement	Multiplying the number of nails by four yields the number of screws.	$4N = S$	$\frac{S}{4} = N$
Operation statement	Six times the number of rhinos is the number of elephants times five.	$5E = 6R$	$\frac{E}{6} = \frac{R}{5}$

The test problems varied along three factors: *phrasing, equation frame format,* and *variable type*. Note that the problems were phrased as relationships among entities on a balanced scale, regardless of the variable type. This was done in order to keep story frames matched across problem types (see the Appendix for sample materials). Table 2 provides examples from each condition and indicates for each condition whether the notational structure of the frame and the story are aligned. The *phrasing* factor manipulated the proximity relations of terms in the story problem through two levels: *direct comparison* and *operation*. In direct comparison items, the critical sentence was of the form "For every A X's, there are B Y's," where A and B were mutually prime number words under 10 and X and Y were two related objects (e.g., pies and cakes). In operation items, the critical sentence was of the form "The number of X's times B is A times the number of Y's." The direct comparison condition makes the pairs A–X and B–Y grammatically proximal (i.e., puts them in the same phrase), while the operation condition makes proximal A–Y and B–X. Furthermore, the ordering of constants and variables is different in the two phrasing conditions, so applying strict translation to operation statements would result in an expression like XB = AY. This property was meant to allow us to distinguish responses generated by a strict left-right transcription strategy; however, even left-right transcription heuristics assume that people accommodate to some degree mathematical norms. As this is not a cleanly distinguishing factor, we will not discuss it further.

The equation frame factor manipulated the proximity relations of the correct responses by requiring the participant to generate either a *multiplication* equation or a *division* equation. Participants were given one of two frames on each problem (see Figure 1) and were instructed to write their answers directly into the equation frame.

The variable type was modified as described above: Each problem either asked participants to write an equation expressing the relationship between the weight of the objects on the scale or asked them to relate the objects' numerosity. Because these properties are inversely related, the correct equation model is reversed for each (to see this, consider that the heavier the object on one side of the scale, the fewer such objects are needed to balance whatever is on the other side of the scale). For the sentence about Janet's cookies above, the equation $7N = 5C$ would be correct if the variables represent counts but would be reversed if they represent weights of individual items.

$$(\quad)\bullet(\quad)=(\quad)\bullet(\quad) \qquad \frac{(\quad)}{(\quad)}=\frac{(\quad)}{(\quad)}$$

Figure 1. Sample equation frames.

Table 2
Manipulations Used in Experiments 2–3

Phrasing	Unit	Frame	Frame–language alignment?
Direct comparison	Number	Multiplication	No
Direct comparison	Number	Division	Yes
Operation or hypothetical comparison	Number	Multiplication	Yes
Operation or hypothetical comparison	Number	Division	No
Direct comparison	Weight	Multiplication	Yes
Direct comparison	Weight	Division	No
Operation or hypothetical comparison	Weight	Multiplication	No
Operation or hypothetical comparison	Weight	Division	Yes

Note. The rightmost column specifies whether the language matched the frame used. Reversal rates are predicted to be higher when frame and language are not aligned.

Predictions. The basic prediction of the notation alignment account is quite simple: Accuracy will be highest when the text places closely together terms that should be placed closely together in the correct equation: This amounts to a full three-way interaction. The dual route account also makes clear predictions. On problems with a multiplication equation frame, participants are likely to engage in left-right transcribing, so accuracy should depend on whether that transcription is correct (transcription yields correct number equations in the operation condition and weight equations in the direct comparison condition). On division frame problems, participants will be unable to engage in left-right transcribing and so will be forced to model. The difficulty of the problem will depend on the difficulty users have in extracting from the text the relevant information. Crucially, on this account comparison phrasing and the variable type act roughly independently, so that these factors should not interact.

Procedure. Written tests were composed containing 16 target items. Target items were separated by a multidigit arithmetic problem that served as a distractor. Equation frames were placed directly below each item. Two target items and two distractors appeared on each page. The order of the conditions was counterbalanced with a Latin square design. The order of the narrative contents was held constant across counterbalancing; story narratives were rearranged to match the particular conditions.

Analysis. In Experiments 2–3 we analyze primarily rates of accuracy and rates of reversal. Because outcomes are bounded averages of binomially distributed variables and also because we anticipated many outcomes to be close to the edges of the range, traditional analyses of variance (ANOVAs) are not appropriate analytical tools. Also, because people occasionally left entries blank and often used scratchwork (see below), Experiment 2 has some missing data. Logistic regressions recode proportions as log-odds and do not assume perfect balance across repeated measures and are therefore conceptually more appropriate tools in this case (Hosmer & Lemeshow, 2000). Throughout the rest of this paper, accuracy and reversal rates are analyzed using mixed-effects logistic regressions with random subject intercepts, which should be robust to all but extreme floor effects. We included all main effects and interactions. These regressions were carried out using the lmer function of the lme4 package in R (Bates & Maechler, 2010). In each case, statistical significance for individual main effects and interactions was determined by a nonparametric permutation bootstrap with 10,000 replications. (Reported p values are in general the proportion of replications with a more extreme measure statistic than the empirical one, in the predicted direction. That is, all p values reflect one-way tests.) In every case, identical significance patterns are obtained with more traditional ANOVA analyses.

In Experiment 2, all results were analyzed using mixed-effects logistic regression models, including main effects of phrasing, frame, and variable type, as well as interactions. We projected, based on pilot work, that many students would produce scratchwork for problems, especially division problems, outside of the problem frame, and so they did: 45% of problems contained some scratchwork. The analyses presented here exclude items in which scratchwork mismatched the given frame. Scratchwork will be taken up again in Experiments 3a and 3b; in the current experiment, note that analyses including these trials yield identical patterns of significance but with generally smaller effects. This is especially true for division-framed problems, in which most scratchwork involved constructing a multiplicative representation. This dominance of scratchwork in division framed problems replicates the bias toward multiplicative equations for algebra users noted by Fisher et al., 2011.

Results

Reversal rates are shown in Figure 2. The primary result of interest is the three-way interaction: The interaction was significant in the predicted direction ($\beta = -21$, $p \sim .001$). Because the notation alignment view uniquely predicts an interaction between variable and phrasing for division problems, we separately explored this two-way interaction as well. Within division problems, the interaction was significant, such that problems that aligned phrasing and equation were reversed less often than problems that did not ($\beta = -2.6$, $p = .011$). Both main effects were also significant: Weight problems led to more reversals than number problems ($\beta = 2.2$, $p = .002$), and operation phrasing led to marginally more errors than direct comparison ($\beta = 1.3$, $p = .028$), suggesting that some problem structures and phrasings were indeed easier to formalize than others.

Within the multiplication problems, once again the predicted interaction held ($\beta = -26$, $p < .0001$), as did both main effects: Weight problems were again reversed more frequently ($\beta = 4.1$, $p < .0001$), as were operation-phrased problems ($\beta = -20$, $p < .0001$). The interaction was even more powerful than for division: Nearly every multiplication frame problem was inverted when the problem was a direct comparison and the variables were numbers and when the variables were weights and the phrasing was operation. These are just the cases for which concrete notational structures misalign with correct multiplication models.

Discussion

This pattern of results closely matched the predictions of the notation alignment view. Reversal rates on problems in which relation binding in English mismatched that of mathematics were very high (averaging 76%); when the notational structures of the two situations aligned, the reversal rate was just 13% on average. The interaction was large and statistically significant in division problems as well as multiplication problems, though the pattern was weaker in division problems overall. The simplest account for this difference is again the multiplication bias observed by Fisher et al. (2011): Many participants may initially have conceived of a multiplication expression and converted it mentally in order to fit it into the required frame.

Experiment 2 provided a strong test of the notation alignment account. However, these problems were clearly quite challenging for our participants. The within-subjects design meant that participants were faced with one difficult problem after another, which

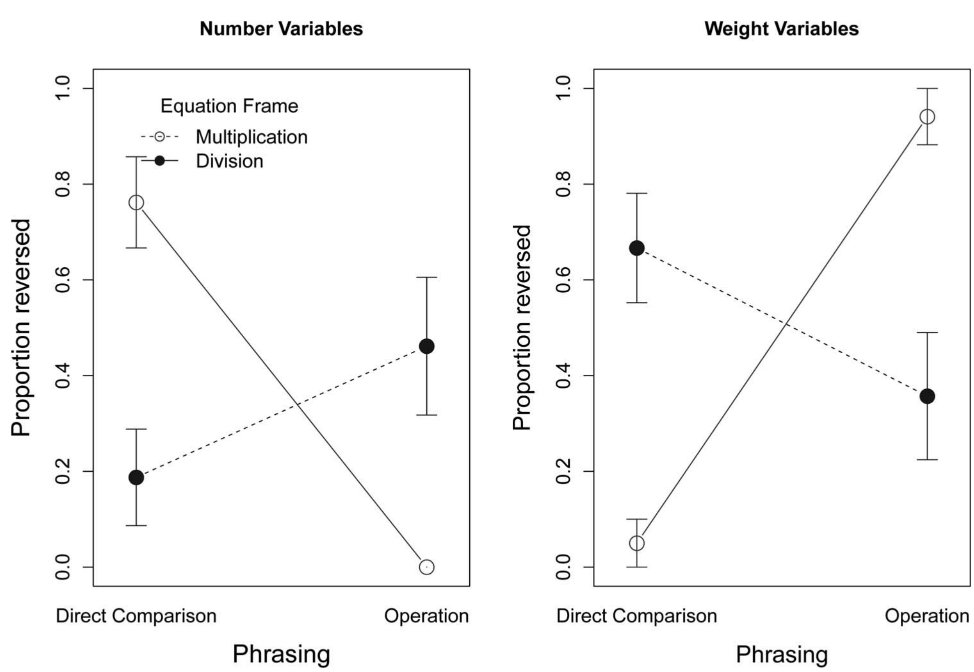

Figure 2. Mean reversal error rate in Experiment 2, across story problem, equation format, and variable type. Error bars represent standard error.

may have affected the strategy they used (cf. Christianson, Mestre, & Luke, 2012). Additionally, the phrasing of the operation statement translates very directly into an algebraic equation and is quite unnatural as an English sentence. These properties may have inclined students away from meaningful semantic modeling.

Experiment 3A explored the same issues as Experiment 2, but two factors were changed. First, a between-participants design was used, in which each participant solved just a single problem. This ruled out the possibility that participants might alter their strategy in the context of a particular (and repetitive) problem set. In addition, Experiment 3A replaced the "operation" phrasing by the less transparently mathematical hypothetical comparison (see Table 1). These two changes largely served to explore the generality of the results already described. In addition, they allow us to evaluate the influence of scratchwork on reversals in more detail.

Experiment 3A

Method

Participants. In this experiment, the large number of participants and the short nature of the task made it difficult to collect data at the home institution of the first and third authors. Participants were recruited from Amazon's Mechanical Turk, an online system for recruiting a scalable workforce to accomplish routine tasks. Although not originally intended as an experiment recruitment site, Mechanical Turk is increasingly used for psychological experimentation. Participants are required by Amazon to be over 18. Participation was further restricted to workers reporting living in the United States; no other restrictions on participation were made. Demographic data suggests that the median U.S. worker on Mechanical Turk (*Turker*) is in the mid-30s, is White, has a bachelor's degree, and is female but that there is substantial diversity in terms of age, education level, and socioeconomic status (Ross, Zaldivar, Irani, & Tomlinson, 2010). Turkers have been found to match college undergraduate populations fairly well on a variety of cognitive tasks (Crump, McDonnell, & Gureckis, 2013; Mason & Suri, 2012).

Participants ($N = 512$) were recruited for this task and were given small monetary compensation. In our sample, the median age was 33 years; of participants, 54% reported being female and 17% self-reported working in a math-related field or having a math-related major in college. Participants who gave answers that were blank or that did not contain relational equations (most of these were numerical equations) were eliminated from consideration, leaving 410 participants in the final analysis.

Design and procedure. Each participant saw a single problem about a person, Olivia, who had gone shopping and purchased hats and bags. Olivia had spent an equal amount of money on each but had purchased four times as many of one as of the other. Three factors were varied. First, problems could be phrased as direct comparisons or hypothetical comparisons. Direct comparisons relate unequal quantities, as in "Olivia bought four bags for every hat." Hypothetical comparisons describe what counterfactuals would make the two sets equal (e.g., "If Olivia had bought four hats for every hat she actually bought, she would have bought as many bags as hats"). Note that in this frame Olivia in fact bought a certain number of hats, but if every hat had instead been four hats, the number of hats would have equaled the number of bags. Second, a problem frame was provided to the participants. As in Experiment 2, this could be either a multiplication or a division equation. Finally, the variable was manipulated such that the requested equation related the number of hats and bags purchased in one level, and the cost of a single hat and bag in the other. Participants were given as much time as they desired to construct their solution. The relative costs and numbers of bags and hats were counterbalanced across participants (the ratio was always 4:1). Because hypothetical comparisons become burdensome with two hypotheticals, we used comparisons with just a single number. Pilot work with this population indicated that with this population reversals would be frequent.

Because participants were recruited online, it was not possible to tell whether they used scratch paper. Our hypothesis, again, is that the default character of multiplication would incline participants to produce a multiplication equation, which could be algebraically converted into a division. Because predictions are reversed for these participants relative to those who directly construct a division frame, it was important to attempt to distinguish them. Two questions were asked to probe this issue. First, participants were directly asked whether they had imagined or written down an equation other than and prior to the one they filled into the frame. Second, participants were asked to briefly describe how they went about solving the problems. The verbal descriptions were separated from condition and other behaviors and were examined by two University of Richmond students ignorant of the hypotheses of the study. Responses were coded based on whether the participant clearly described the (physical or mental) construction of an equation prior to the one in the frame. For example, "It was 4 hats to every bag so $4H = B$, but the problem wanted a division solution, so I divided by sides by 4 and came up with the ratio," "Four hats = Bags => $4H = B$ => $H = B/4$," and "I just thought $4B = H$" were all coded as initial multiplication equations; verbal descriptions could be coded as *multiplication, division,* or *unclear*. The two raters rated all descriptions. Their level of agreement was moderate: Coders agreed on 76% of ratings (Cohen's $\kappa = 0.56$), indicating only a moderate level of agreement between raters. The first rater was not available for consultation, so discrepancies were resolved through discussion between the second rater and the first author using just the verbal descriptions. Finally, the response was coded as *scratchwork* if a participant reported picturing or writing a different equation, or if the participant's description was coded as using an initial equation that mismatched the provided frame.

Predictions. In the notation alignment view, problems will be easy when the text aligns well with the correct response for the variables given and the frame initially used. For instance, with the multiplication frame and the cost variable (the traditional students-and-professors case of Clement et al., 1981), hypothetical comparisons should be formalized more readily than direct comparisons, because their concrete and semantic properties align well with the same correct response. For division frames or cost variables, the direct comparisons should lead to fewer reversals than hypothetical comparisons, but when the frame is division and the variable is cost, the hypothetical comparisons should again be easier. Finally, all of these patterns should hold only when an initial mismatching frame is not used; the relations should be reversed when the participant initially constructs a mental model different from the provided form. Thus, the notation alignment view predicts a four-way interaction among frame, phrasing, variable, and the use of scratchwork.[1]

The dual systems account does not make a strong prediction about accuracy in most of these cases. In direct comparison statements, some reasoners will be lured into making left-right transcriptions with multiplication frames (or imagined multiplicative equations), while others will engage in modeling. In the hypothetical comparisons, a process of strict left-right transcription is blocked by the phrasing (which has too many instances of some variables as well as some ordering problems), but students who engage in modeling will arguably have a harder time doing so, because the relations involved are less transparent in this phrasing. Thus, which phrasing is associated with higher error rates will depend on whether more students are lured into transcription in the direct comparison or more students make modeling errors in the hypothetical. In the division case, accuracy should be no lower (and presumably higher) in the more easily modeled direct comparison case and should not depend on variable type except perhaps as a main effect (if it is for some reason easier to model costs or numbers).

Results

Reversals were analyzed using nested mixed-effects logistic regression models, including main effects of phrasing, frame, variable type, and scratchwork, as well as all interactions to estimate reversal rates. Reversal rates are shown in Figure 3. The four-way interaction was marginally significant ($\beta = -2.8$, $p = .07$; notice that, like all comparisons reported in this paper, this was effectively a one-way comparison) and in the predicted direction. When participants reported directly writing the solution into the frame without an intervening visualization of an equation, reversals were as predicted by the notation alignment account. When visualization was reported, the interaction was inverted for division-framed problems; multiplication problems were less affected by reported scratchwork. According to planned comparisons within each problem type, the three-way interaction was significant for divisions ($\beta = -3.03$, $p = .01$) but not for multiplications ($\beta = -0.16$, $p = .95$). Within divisions, there was a trend in the same direction as Experiment 2 ($\beta = -1.40$, $p = .09$) for trials that were reported as solved directly, while it was reversed for those scratchwork ($\beta = 1.1$, $p = .02$). Within multiplications, the pattern of Experiment 2 held both when problems were solved directly ($\beta = 3.03$, $p < .001$) and when they were not ($\beta = 3.32$, $p = .002$).

Discussion

In each case, the predictions of the notation alignment account are borne out. Systematically, across a wide range of problem types and phrasings, the organization of the problem description shapes the structure of typical errors in equation production. These patterns are parsimoniously described by the idea that equations are constructed and identified via relational alignment processes that include the structure of notational representations.

The marginally significant four-way interaction depended on two assumptions: (a) Alignment of concrete notational structures matters across frames, phrasing structures, and problems. (b) Multiplication is a default format, so that scratchwork was more likely to convert a division problem into a multiplication than vice versa. Our detection of the four-way interaction was limited by our ability to resolve visualizations. Even written scratchwork is probably only a mediocre proxy for initial visualizations; it is likely that our methodology treated a few people who were solving the problem directly as having written scratchwork and missed many people who imagined an equation but failed to report having done so. This possibility is made especially likely by the fact that interrater agreement was only moderate. Furthermore, the primary four-way interaction was only marginally significant. On the other hand, every specific three-way and two-way comparison matched the predicted direction (and replicated Experiment 2), giving us some confidence in the predicted interaction.

The interaction of primary interest in Experiment 3A was only marginally significant, so we decided to investigate the patterns further with a follow-up experiment. In Experiment 3A, by hypothesis, the presented frame does not really matter to the outcomes; what principally drives performance is the analogical alignment between the problem and the first equation frame the participant creates or imagines. Most participants are highly skilled at the simple algebraic transformations needed to convert between division and multiplication formats. The division frame encourages use of division but otherwise does not particularly affect the outcome of the result. In Experiment 3B, we simplified the intended interaction by consistently presenting a single frame type and explicitly asking participants to report whether their initial frame was a multiplication or division (or something else), eliminating the need for coding brief verbal descriptions. On the basis of Experiment 3A, we predict a three-way interaction among variable type, participant-chosen frame, and phrasing, such that problems will be reversed less often when the notationally aligned response produces the correct answer.

Experiment 3B

Method

Participants. Participants ($N = 320$) were recruited from Amazon's Mechanical Turk. We selected this number based on the estimated power of the predicted interaction, using the data from Experiment 3A and pilot studies.

Design and procedure. The design was similar to Experiment 3A, with two differences. First, the multiplication frame problems were dropped in order to focus on division problems. All participants filled in the division frame as their response. Second, the strategy probe was more direct: Participants were asked to report whether the first equation they wrote, imagined, or entered was a multiplication or a division equation. Examples of each were provided. Participants were allowed an *other* option, in which case they were asked to explain their response. 82 (25%) of the responses were *other*. Inspection of the strategy explanations in the *other* reports indicated that nearly all reported strategies compatible with filling in a frame of one type or another. Including these reports did not affect the results reported here, but for these analyses those results are excluded.

[1] We thank Tom Carr for suggesting this analysis.

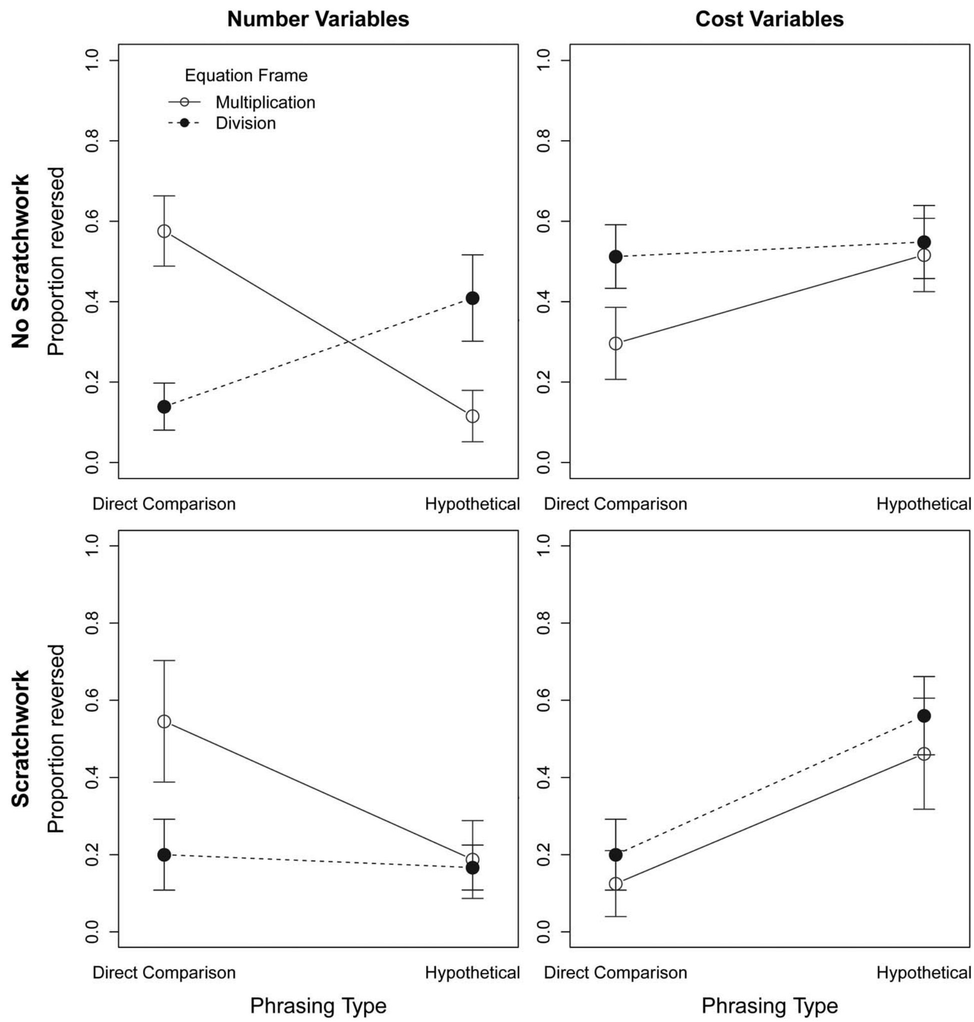

Figure 3. Reversal error pattern in Experiment 3A. The error bars represent within-bin standard errors.

Results and Discussion

Reversal rates are shown in Figure 4. Notice that the bins in this graph are not equal, because strategies followed participants' reports: In the analyzed set, 158 participants reported using division frames, and 80 participants reported using multiplication frames. The primary interaction of interest is the three-way interaction among problem phrasing, reported initial frame, and variable type. This interaction was significant ($\beta = -4.5$, $p < .0001$), in the predicted direction. Within the group reporting using division, we found the predicted interaction favoring aligning variable type and phrasing ($\beta = -2.5, p \sim .001$). For multiplication users, for whom the alignments were inverted, there was a marginal trend in the inverse (predicted) direction ($\beta = 2.1$, $p \sim .07$).

The pattern of results matched those of Experiment 3A, and the main interaction was statistically significant. Overall, the impact was quite large: The reversal rate when aligning correct syntax across representations was 0.26, while that of more difficult inverted problems was 0.56. Overall, the reversal rate was not lower among division (0.41) than multiplication productions (0.38). The reversal rate was lower among the more easily transcribed direct comparisons (0.31) than among hypothetically phrased problems (0.49, $p \sim 0.01$). If anything, division problems were more sensitive to notational alignment than were multiplication problems. Although divisions may well encourage richer modeling than do standard multiplication equations, this does not eliminate the important role played by notational alignment.

General Discussion

Four experiments matched predictions made by the notation alignment account. In equations for which two variable orders were equally correct and conventional, undergraduate physics majors were much more likely to construct an equation that matched the order of introduction in the story they were given, even when that equation had to be retrieved from memory. People apparently engaged in situation modeling and story understanding nevertheless show bias in favor of story order, demonstrating that extra heuristics and conceptual confusions are not necessary to account for reversals. Furthermore, across a wide range of equation structures and story formats, including those previously hypothesized to resist heuristic transcription, to encourage modeling, or to support

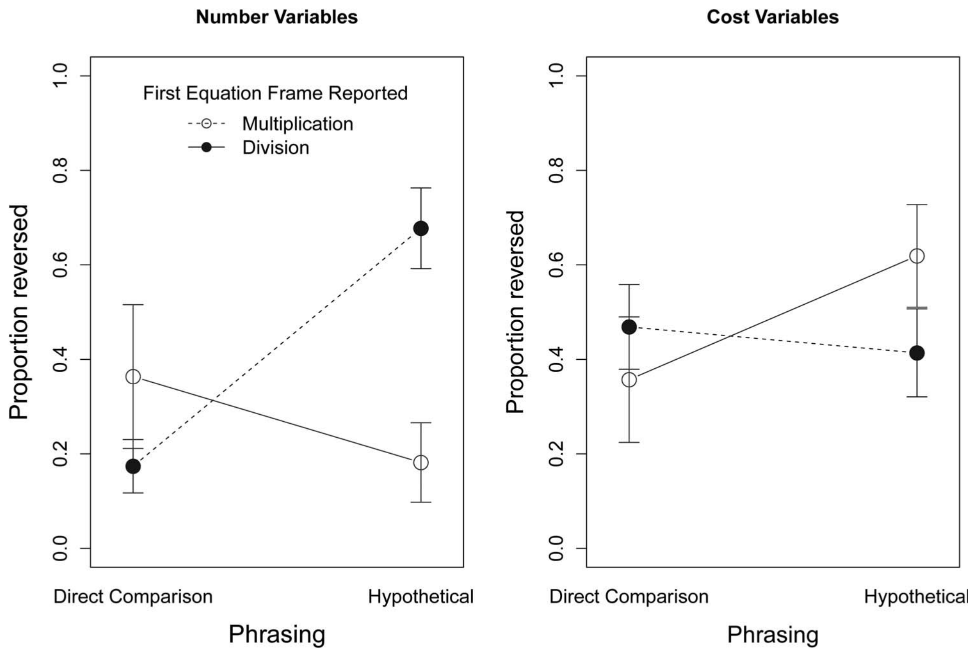

Figure 4. Reversal error pattern in Experiment 3B. The error bars represent within-bin standard errors.

operational interpretations of symbols, reversal rates were predicted by the alignment between the notational structure of provided or chosen equation formats and source story problems. These results are all predicted if we start by assuming structure mapping as a core process in the coordination of diagrams, equations, and other external aids to thought (Gentner, 1983; Martin & Bassok, 2005; Nersessian 1992).

Participants constructing division equations were affected by notation relations in problem statements, just as are people writing the more common multiplication equations. As reported by Fisher et al. (2011), students do seem to show a strong preference for multiplications in algebraic statements over divisions, and this preference does seem to interfere with success. However, the benefits of division equations do not lie in their blocking the use of surface details but precisely in the fact that the apparent structure of divisions is aligned with that of the most natural English-language relation statements.

All this is not to suggest that no student ever engages in pure transcription, and it is certainly not meant to suggest that no student ever misunderstands the difference between variables and units. However, these errors appear not to be unique or even frequent causes of reversals in relational equation construction. The relative prominence of different strategies of equation generation is important for understanding mathematical practice and mathematics education, though it is clearly one of degree rather than of kind. Misconstruing student errors may be counterproductive both to helping students overcome these patterns of errors and to producing comprehensive theoretical accounts of the role of formal notations in reasoning.

Reversal errors are present in nearly all cases, including those that are phrased so as to align notation structures. That fact has proved difficult for the dual strategies account to explain (Cohen & Kanim, 2005; cf. Christianson et al., 2012). In our view, part of the difficulty stems from an inappropriate focus on finding the source of errors: Dual route accounts treat correct behavior as an uninteresting default (see also MacGregor & Stacey, 1993). A more appropriate conceptual approach to the problem of equation production is to ask how students succeed at all. Although the dual route account generally assumes that understanding of algebraic syntax is fairly robust, substantial evidence suggests that it is not. For instance, students are often more successful at solving problems phrased in natural language than in algebraic syntax (Koedinger, Alibali, & Nathan, 2008; Koedinger & Nathan, 2004). Despite the difficulty students have in forming symbolic equations to model multiplicative relations, they do not seem to have trouble interpreting the natural-language expressions: Around 90% are able to correctly report that there are more students than professors (Martin & Bassok, 2005; Wollman, 1983) in the standard phrasing; when the symbolic relation is replaced by numerical relations, college-level students have little difficulty solving structurally identical problems (Wollman, 1983).

In the alignment view espoused here and in Fisher et al. (2011), the successful procedure involves two (nonsequential) tasks: selecting an appropriate equation frame and aligning selected frames with the source situation. If a reasoner succeeds in selecting an appropriate frame for a multiplicative relational equation (regardless of whether the operation chosen is multiplication or division), in the absence of alignable information the reasoner will have an equal chance of producing correct and inverted equations. If this is a core component of the process of formalization, then we cannot expect to eliminate it through direct training in algebraic syntax. Indeed, we conducted a pilot study with graduate students in physics attending a highly ranked university, asking them to solve two-variable problems with count variables like those discussed here (*There are five rhinos for every six elephants*). With the standard direct comparison and multiplicative frame, graduate students in physics gave reversed equations 48% of the time (95% CI [0.24, 0.71]). That graduate students in physics who make relatively few other errors and are surely free of most conceptual errors about mathematical syntax nevertheless reverse equations at a high rate suggests a central role for notational structures in alignment.

The better question to ask is not the source of errors, but what sorts of activities influence reasoners away from this base rate. In each of our tasks, participants were more likely overall to produce correct than incorrect responses. One possibility is that semantic knowledge contributed to overcome incorrect mappings. Another is that success was driven by simple but useful strategies such as checking answers by substituting plausible numbers, or verifying relative magnitude relations among the quantities. In the context of alignment, these can be seen as probing the dynamics of an analogy; that is, making sure that the text and equation transform isomorphically. In any event, squarely framing the problem of equation modeling as one of explaining success, not failure, seems likely to facilitate progress in deriving successful training procedures and in constructing productive models of reasoner behavior.

Our capacity for wholly abstract relational reasoning may be strikingly limited. Algebra and its accompanying notation are a paradigmatic case of purely symbolic thought. That experienced users of algebra rely on concrete physical structure suggests the interpretation that purely symbolic thought is itself largely achieved not through complex abstract internal resources but through the cooption—in this case via external formal notations—of resources typically devoted to representing concrete relations and features. On this view, notation does not mirror thought. Instead, notation gives us something to think about.

References

Anderson, J. R. (2005). Human symbol manipulation within an integrated cognitive architecture. *Cognitive Science, 29,* 313–341. doi:10.1207/s15516709cog0000_22

Bassok, M., Chase, V. M., & Martin, S. A. (1998). Adding apples and oranges: Alignment of semantic and formal knowledge. *Cognitive Psychology, 35,* 99–134. doi:10.1006/cogp.1998.0675

Bassok, M., Wu, L., & Olseth, L. K. (1995). Judging a book by its cover: Interpretative effects of content on problem solving transfer. *Memory & Cognition, 23,* 354–367.

Bates, D., & Maechler, M. (2010). lme4: Linear mixed-effects models using S4 classes (R package version 0.999375–35) [Computer software]. Retrieved from http://CRAN.R-project.org/package=lme4

Bever, T. G., Jandreau, S., Burwell, R., Kaplan, R., & Zaenan, A. (1990). Spacing printed text to isolate major phrases improves readability. *Visible Language, 25,* 74–87.

Carlson-Radvansky, L. A., & Logan, G. D. (1997). The influence of reference frame selection on spatial template construction. *Journal of Memory and Language, 37,* 411–437. doi:10.1006/jmla.1997.2519

Chatterjee, A., Southwood, M. H., & Basilico, D. (1999). Verbs, events and spatial representations. *Neuropsychologia, 37,* 395–402. doi:10.1016/S0028-3932(98)00108-0

Christianson, K., Mestre, J. P., & Luke, S. G. (2012). Practice makes (nearly) perfect: Solving “students-and-professors”-type algebra word problems. *Applied Cognitive Psychology, 26,* 810–822. doi:10.1002/acp.2863

Clement, J. (1982). Algebra word problem solutions: Thought processes underlying a common misconception. *Journal for Research in Mathematics Education, 13,* 16–30. doi:10.2307/748434

Clement, J., Lochhead, J., & Monk, G. S. (1981). Translation difficulties in learning mathematics. *American Mathematical Monthly, 88,* 286–290. doi:10.2307/2320560

Cohen, E., & Kanim, S. E. (2005). Factors influencing the algebra “reversal error.” *American Journal of Physics, 73,* 1072–1078. doi:10.1119/1.2063048

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon’s Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE, 8,* e57410. doi:10.1371/journal.pone.0057410

Dantzig, T. (1930). *Number: The language of science. A critical survey written for the cultured non-mathematician*. New York, NY: Macmillan.

Fisher, K. J., Borchert, K., & Bassok, M. (2011). Following the standard form: Effects of equation format on algebraic modeling. *Memory & Cognition, 39,* 502–515. doi:10.3758/s13421-010-0031-6

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7,* 155–170. doi:10.1207/s15516709cog0702_3

Goldstone, R. L., Landy, D., & Son, J. Y. (2010). The education of perception. *Topics in Cognitive Science, 2,* 265–284. doi:10.1111/j.1756-8765.2009.01055.x

Hegarty, M., Canham, M. S., & Fabrikant, S. I. (2010). Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 37–53. doi:10.1037/a0017683

Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology, 87,* 18–32. doi:10.1037/0022-0663.87.1.18

Hinsley, D., Hayes, J. R., & Simon, H. A. (1977). From words to equations. In P. Carpenter & M. Just (Eds.), *Cognitive processes in comprehension* (pp. 89–106). Hillsdale, NJ: Erlbaum.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York, NY: Wiley.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review, 104,* 427–466. doi:10.1037/0033-295X.104.3.427

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York, NY: Basic Books.

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2008, April 25). The advantage of abstract examples in learning math. *Science, 320,* 454–455. doi:10.1126/science.1154659

Kirshner, D. (1989). The visual syntax of algebra. *Journal for Research in Mathematics Education, 20,* 274–287. doi:10.2307/749516

Kirshner, D., & Awtry, T. (2004). Visual salience of algebraic transformations. *Journal for Research in Mathematics Education, 35,* 224–257. doi:10.2307/30034809

Koedinger, K. R., Alibali, M. W., & Nathan, M. J. (2008). Trade-offs between grounded and abstract representations: Evidence from algebra problem solving. *Cognitive Science, 32,* 366–397. doi:10.1080/03640210701863933

Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *Journal of the Learning Sciences, 13,* 129–164. doi:10.1207/s15327809jls1302_1

Kranjec, A., Lehet, M., Bromberger, B., & Chatterjee, A. (2010). A sinister bias for calling fouls in soccer. *PLoS ONE, 5,* e11667. doi:10.1371/journal.pone.0011667

Landy, D., & Goldstone, R. L. (2007a). Formal notations are diagrams: Evidence from a production task. *Memory & Cognition, 35,* 2033–2040. doi:10.3758/BF03192935

Landy, D., & Goldstone, R. L. (2007b). How abstract is symbolic thought? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 720–733. doi:10.1037/0278-7393.33.4.720

Maass, A., & Russo, A. (2003). Directional bias in the mental representation of spatial events. *Psychological Science, 14,* 296–301. doi:10.1111/1467-9280.14421

MacGregor, M., & Stacey, K. (1993). Cognitive models underlying students' formulation of simple linear equations. *Journal for Research in Mathematics Education, 24,* 217–232. doi:10.2307/749345

Malle, G. (1993). *Didaktische Probleme der elementaren Algebra* [Didactical problems of elementary algebra]. Wiesbaden, Germany: Vieweg.

Marquis, J. (1988). Common mistakes in algebra. In A. F. Coxford & A. P. Shulte (Eds.), *The ideas of algebra, K-12: 1988 yearbook* (pp. 204–205). Reston, VA: National Council of Teachers of Mathematics.

Martin, S. A., & Bassok, M. (2005). Effects of semantic cues on mathematical modeling: Evidence from word-problem solving and equation construction tasks. *Memory & Cognition, 33,* 471–478. doi:10.3758/BF03193064

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods, 44,* 1–23. doi:10.3758/s13428-011-0124-6

Mestre, J. P., & Lochhead, J. (1983). The variable-reversal error among five cultural groups. In J. Bergeron & N. Herscovics (Eds.), *Proceedings of the Fifth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 180–188). Montreal, Ontario, Canada: International Group for the Psychology of Mathematics Education.

Mochon, D., & Sloman, S. A. (2004). Causal models frame interpretations of mathematical equations. *Psychonomic Bulletin & Review, 11,* 1099–1104. doi:10.3758/BF03196743

Nersessian, N. J. (1992). How do scientists think? Capturing the dynamics of conceptual change in science. In R. N. Giere (Ed.), *Cognitive models of science* (pp. 3–45). Minneapolis: University of Minnesota Press.

Novick, L. R., & Catley, K. M. (2007). Understanding phylogenies in biology: The influence of a Gestalt perceptual principle. *Journal of Experimental Psychology: Applied, 13,* 197–223. doi:10.1037/1076-898X.13.4.197

Paige, J. M., & Simon, H. A. (1966). Cognitive processes in solving algebra word problems. In B. Kleinmuntz (Ed.), *Problem solving: Research, method, and theory* (pp. 51–119). New York, NY: Wiley

Patsenko, E. G., & Altmann, A. M. (2010). How planful is routine behavior? A selective-attention model of performance in the Tower of Hanoi. *Journal of Experimental Psychology: General, 139,* 95–116. doi:10.1037/a0018268

Reed, S. K. (1987). A structure-mapping model for word problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 124–139. doi:10.1037/0278-7393.13.1.124

Ross, J., Zaldivar, A., Irani, L., & Tomlinson, B. (2010). *Who are the Turkers? Worker demographics in Amazon Mechanical Turk*. Retrieved from http://www.ics.uci.edu/~jwross/pubs/SocialCode-2009-01.pdf

Sherin, B. L. (2001). How students understand physics equations. *Cognition and Instruction, 19,* 479–541. doi:10.1207/S1532690XCI1904_3

Weaver, C. A., III, & Kintsch, W. (1992). Enhancing students' comprehension of the conceptual structure of algebra word problems. *Journal of Educational Psychology, 84,* 419–428. doi:10.1037/0022-0663.84.4.419

Wollman, W. (1983). Determining the sources of error in a translation from sentence to equation. *Journal for Research in Mathematics Education, 14,* 169–181. doi:10.2307/748380

Appendix

Sample Materials

Experiment 1

There were two stories, altering the order of introduction of the elements and which mass was in motion.

Single Asteroid First

Imagine an asteroid, with mass m. This asteroid flies through space, occasionally coming close enough to other asteroids for gravity between them to become nonnegligible. We'll consider several such instances, and in each case use Newton's law of universal gravitation to find the attractive force between the asteroid and its current nearest neighbor.

Several Asteroids First

Imagine a group of asteroids, with masses m_1, m_2, and m_3. These asteroids fly through space, and around a certain time come close enough to another asteroid (mass $= m$) for gravity to become nonnegligible. We'll consider each asteroid in turn, and in each case use Newton's law of universal gravitation to find the attractive force between each moving asteroid and the stable asteroid.

Experiment 2

Questions from one condition are presented. For other conditions, the text was inverted, but the narratives were fixed. An equation frame was presented beside each figure. The frames used are shown in Figure 1.

1. Janet is baking cookies. She needs to have the same weight of both chocolate chips and nuts. Measuring these ingredients on a balance, she notices that for every seven chocolate chips on one side of the balanced scale, there are five nuts on the other side. Using C to represent the **weight** of one chocolate chip, and N to represent the **weight** of one nut, fill in the equation below so that it correctly expresses the relationship between the chocolate chips and nuts when the scale is balanced.

2. A bakery sells cakes and pies by weight. To determine what price it should charge for each dessert, it compares the weights of

(*Appendix continues*)

cakes and pies using a scale. The number of cakes times four on one side of the balanced scale is five times the number of pies on the other side. Using C to represent the **number** of cakes and P to represent the **number** of pies, fill in the equation below so that it correctly expresses the relationship between the cakes and pies when the scale is balanced.

3. A veterinarian is comparing the weight of kittens to the weight of puppies. The number of kittens times three on one side of the balanced scale is two times the number of puppies on the other side. Using K to represent the **weight** of one kitten and P to represent the **weight** of one puppy, fill in the equation below so that it correctly expresses the relationship between the kittens and puppies when the scale is balanced.

4. A nutritionist explores the difference in weight between boys and girls using an enormous balance. She places boys on one side of a scale and girls on the other. For every nine boys on one side of the balanced scale, there are seven girls on the other side. Using B to represent the **number** of boys and G to represent the **number** of girls, fill in the equation below so that it correctly expresses the relationship between the boys and girls when the scale is balanced.

5. A University of Richmond English professor assigns readings based on how heavy different books are. Before class one day, when balancing a scale in his office, he discovers that the number of textbooks times three on one side of the balanced scale is five times the number of fiction books on the other side. Using T to represent the **weight** of one textbook and F to represent the **weight** of one fiction book, fill in the equation below so that it correctly expresses the relationship between the textbooks and fiction books when the scale is balanced.

6. A jeweler uses a scale to compare the weight of emeralds and the weight of rubies. For every three emeralds on one side of the balanced scale, there are two rubies on the other side. Using E to represent the **number** of emeralds and R to represent the **number** of rubies, fill in the equation below so that it correctly expresses the relationship between the emeralds and rubies when the scale is balanced.

7. A hardware store owner places screws on one side of a scale and nails on the other side. She notices that for every 11 screws on one side of the balanced scale, there are 13 nails on the other side. Using S to represent the **weight** of one screw and N to represent the **weight** of one nail, fill in the equation below so that it correctly expresses the relationship between the screws and nails when the scale is balanced.

8. Bob has a scale. There are some deegers on one side and some koozles on the other side of a balanced scale. The number of deegers times eight on one side of the balanced scale is five times the number of koozles on the other side. Using D to represent the **number** of deegers and K to represent the **number** of koozles, fill in the equation below so that it correctly expresses the relationship between the deegers and koozles when the scale is balanced.

Experiments 3A and 3B

Two conditions are presented; as described in the main text, variable and comparison type were crossed in the actual study. Furthermore, the role of "bags" and "hats" were counterbalanced throughout the description and instructions.

Instructions

Please read the following story, and follow the instructions at its end. Please don't write anything down; solve the problem in your head, and write the answer directly into the form.

Hypothetical Comparison and Cost Condition

Olivia went shopping for clothes. She found a great sale and bought many hats and bags. At the sale, bags cost a certain price, and hats cost a different price. As she traveled home, she thought about what she had bought. Olivia noticed that she spent the same amount on hats as she did on bags.

Olivia loves hats but already has too many. She had to deliberately stop herself from buying many more than she did. Olivia noticed that if she had bought four hats for every hat she actually bought, she would have bought as many bags as hats.

Use *H* to represent the COST OF ONE HAT at the sale Olivia found, *B* to represent the COST OF ONE BAG, and the number 4 to fill in the equation below so that it reflects the relationship between hats and bags.

Direct Comparison and Number Condition

Olivia went shopping for clothes. She found a great sale and bought many hats and bags. At the sale, bags cost a certain price, and hats cost a different price. As she traveled home, she thought about what she had bought. Olivia noticed that she spent the same amount on hats as she did on bags.

Olivia loves bags but already has too many. She had to deliberately stop herself from buying many more than she did. Olivia noticed that she bought four hats for every bag.

Use *H* to represent the NUMBER OF HATS Olivia actually bought, *B* to represent the NUMBER OF BAGS she actually bought, and the number 4 to fill in the equation below so that it reflects the relationship between hats and bags.

Received August 23, 2012
Revision received February 19, 2014
Accepted February 20, 2014 ■