

Causal models frame interpretation of mathematical equations

DANIEL MOCHON and STEVEN A. SLOMAN
Brown University, Providence, Rhode Island

We offer evidence that people can construe mathematical relations as causal. The studies show that people can select the causal versions of equations and that their selections predict both what they consider most understandable and how they expect variables to influence one another. When asked to write down equations, people have a strong preference for the version that matches their causal model. Causal models serve to structure equations by determining the preferred order of variables: Causes should be on one side of an equality, and a single effect should appear on the other.

Mathematical equations provide one means by which to represent relations in the world. Mathematical representation has several advantages, including ease of derivation and calculation. Equations can also provide a concise and exact representation while revealing the variables that are and are not relevant to a phenomenon. One mathematical property of equations is that they are symmetric, in the sense that any variable can appear on either side of an equality; $X = YZ$ means the same thing as $Y = X/Z$. The hypothesis we test in this article is that equations have a conceptual function beyond their mathematical one; they provide a tool for thinking. Specifically, we demonstrate that, in the causal domain, equations are not understood symmetrically but, rather, in correspondence with the conceiver's causal model.

Learning to solve word problems in physics and mathematics has been examined in many studies (e.g., Hall, Kibler, Wenger, & Truxaw, 1989; Paige & Simon, 1966). Theorists have recognized that general, schematic knowledge is critical to the problem-solving process (Hinsley, Hayes, & Simon, 1977; Nathan, Kintsch, & Young, 1992; Priest & Lindsay, 1992). But theories have not been directly concerned with how equations themselves are understood (Sherin, 2001), and therefore, symmetry has not been questioned. Rather, problem solving has been analyzed as a process of search through a space of equations where the learner begins with what is given in the problem and tries to find the problem's solution state. Even Priest and Lindsay, who posit a set of schema as the building blocks of problem solving, focus on the process of problem solving without making any claim about how equations per se are understood.

On all analyses, equations must be constructed or retrieved from memory. For example, equations might be embedded in schema whose activation elicits the equation (cf. Chi, Feltovich, & Glaser, 1981; Larkin, 1983). We propose that the equation emerges with a preferred form. Sherin (2001) suggested that equations are not mere collections of symbols but are understood to express meaning; their form reflects the conceptual structure being represented. Likewise, we will argue that formulas are understood and written down to correspond to conceptual structure in the form of causal models.

Sherin (2001) has posited conceptual schema associated with various arithmetic operations. Our focus is both more specific and more general. It is more specific in that it concerns only the causal domain. We will argue that when people have beliefs about the causal mechanisms that relate the variables in an equation, they prefer the equation to take the form that corresponds to their causal model. In particular, given the option, people will choose an equational form that corresponds to that of structural equations or causal graphs: a single effect on one side, all causes on the other. Our suggestion is more general than Sherin's in not being limited to any specific mathematical operation. Whatever the functional relation, we posit that people will tend to isolate effects, and not causes.

Causal representation requires asymmetry; causes and effects are distinct. For a given causal relation, a variable can play only one role. Causal asymmetry is important because it affords action and counterfactual inference. The representation of causal structure is essential for intervening in the world physically or in imagination. Such a representation can support causal inference and learning in the form of graphical probability models (Pearl, 2000; Spirtes, Glymour, & Scheines, 1993). In general, such a representation is a model of the world that takes causes as inputs and produces effects as outputs and, thus, affords simulation of events.

Evidence for the psychological reality of causal models has been reviewed in Glymour (2001), Sloman and

This work was funded by NASA Grant NCC2-1217. We thank Bruce Sherin for editorial comments on a previous version and David Sobel, Sean Stromsten, and especially Dave Lagnado and Michael Waldmann for their help on this project. Correspondence concerning this article should be addressed to S. A. Sloman, Cognitive and Linguistic Sciences, Brown University, Box 1978, Providence, RI 02912 (e-mail: steven_sloman@brown.edu).

Lagnado (2004), and Waldmann (1996). Their relevance to the conceptualizing of equations is suggested by several findings. First, teaching causal structure often helps form a deeper understanding (White, 1993). Second, causal models are critical to physics comprehension (diSessa, 1993; Hunt & Minstrell, 1994; Reif & Allen, 1992). Third, people impose a causal frame even when it distorts a representation. For example, people assert a unidirectional causal relation between variables that interact dynamically. The circuit between a battery and a light is often incorrectly understood as unidirectional, with energy flowing from the battery to the light (Andersson, 1986; Driver, Guesne, & Tiberghien, 1993; Reiner, Slotta, Chi, & Resnick, 2000).

How do these two forms of representation, causal and mathematical, relate? Following Pearl (2000), we propose that causal models, when available, determine the favored form of mathematical equations. Equations should be preferred and comprehended more easily if the order of their variables corresponds to the form of a causal representation: Only one of the variables should be an effect and it should be on one side of the equal sign; the remaining variables (causes) should be on the other. This is the form of the structural equations that are central to graphical causal modeling (Pearl, 2000). Current supporting evidence is that people consider the left side of an equation to be the outcome of a change in something on the right (Sherin, 2001).

The hypothesis predicts that, in the case of a three-variable equation that matches a single causal model, the

variable that is an effect of the other two should be written on one side and the two causes on the other. For example, the equation that relates pressure (P), weight (W), and area (A) can be written three different ways: $P = W/A$, $A = W/P$, and $W = AP$. $P = W/A$ should seem most intuitive, because the most natural naive causal model represents pressure as the effect, not as the cause, of weight and area. This is suggested by the intuition that if weight were increased, pressure would increase. The equation that solves for A is not as understandable; the intuition that increasing weight would increase area is weaker. Due to the convention for writing equations, the effect is usually expected to be written on the left and the causes on the right, but the critical prediction is merely that the effect is isolated.

In sum, we posit a correspondence between naive causal structure and the form in which equations are most easily understood; causal structure brings an asymmetry to the representation. In four studies, we tested whether people have a preference for the version of a set of equivalent equations that matches causal understanding. We used 16 equations from various fields, including physics, biology, and economics (see Table 1). They were grouped into three classes. The first type (*causal*) were equations with a clear causal model (e.g., $P = W/A$). These represent phenomena that are familiar, so most people understand what causes what. We assumed that the causal models of these relations would be consistent across our population of Brown and Yale University undergraduates. The second type (*unclear*) were equations that rep-

Table 1
Proportion Choosing One Form of Each Equation in Studies 1 and 2, Modal Model Constructed, and Percentage of Participants Indicating Links From Both Causes to Effect in Study 3

	Study 1	Study 2	Model	% With Two Causes
Causal				
1. behavior potential = expectancy of the reinforcement * value of reinforcement	.82	.70	A	67
2. grade = understanding/difficulty of the test	.65	.55	B	89
3. growth rate = birth rate - death rate	.94	.75	A	89
4. neuron firing rate = amount of excitatory enzyme/amount of inhibitory enzyme	.65	.35	A	41
5. price = money demand/money supply	.76	.70	C	94
6. growth of crops = amount of rain/bags of pesticide	.18	.75	A	89
7. force = mass * acceleration	.76	.70	C	78
8. current = voltage/resistance	.53	.75	C	67
9. pressure = weight/area	.76	.40	A	67
Unclear				
10. speed = Planck's constant/mass * wavelength	.29	.55	N/A	N/A
11. thermal efficiency = work done during one cycle/heat added during one cycle	.35	.45	A	44
12. heat capacity = change in heat/change in temperature	.29	.60	C	44
13. internal column load = axial deformation * area * young's modulus/length	.35	.20	N/A	N/A
Noncausal				
14. $y = x * z$.24	.30	other	33
15. Kelvin = Celsius degrees + 273	.35	.40	N/A	N/A
16. length of the hypotenuse ² = length of one side ² + length of the other side ²	.53	.90	other	94

Note—For the causal equations, the causal form is shown. For the noncausal and unclear equations, the modal choice from Study 1 is shown. The participants were presented with equations that related one-letter variables, with an explanation of what each variable meant above the equations. Equations 2, 4, and 6 were fabricated for these studies. All other equations were found in textbooks. For Studies 1 and 2, the values represent the modal choice for each equation. If the modal choice indicated no preference, the value for the most frequently chosen form is shown. This sometimes leads to a low proportion as the modal choice (e.g., Study 1, Equation 6). Values in bold are for equations whose form corresponds to the form written in the table.

resent phenomena that are not very familiar but for which people have uncertain causal beliefs. Therefore, we expected different people to have different causal models. One example relates thermal efficiency (η), work done during one cycle (W), and heat added during one cycle (Q): $\eta = W/Q$. People likely have weak and variable intuitions about how thermal efficiency relates to work. The third type (*noncausal*) were equations with no causal interpretation. They represent geometric relations or scale transformations. An example is the equation that gives the relation between temperature in Kelvin and Celsius: $K = ^\circ C + 273$.

STUDY 1

In the first study, we tested which version of the equations participants would choose when asked for the causal version of each from among its possible mathematically equivalent forms. If equations are better understood to the extent they are related to causal models, people should be able to recognize the versions of equations that match these models. Therefore, we predicted that the different types of equations would show different patterns. For the causal equations, most participants should choose the version that we predicted to be causal. For the unclear equations, we would expect a uniform distribution across the different causal versions. For the noncausal equations, most people should choose the “none of them are causal” option.

Method

Participants. Eighteen undergraduates from an advanced cognitive science laboratory class at Brown University were tested. The data from one of the participants was excluded because he chose all of the options, including the contradictory “none is causal.”

Procedure. All the participants received a questionnaire containing, for each equation, all possible forms isolating a single variable, as well as a “none of them is causal” option. They were asked to choose the version of each equation they considered causal, defined as “the version of the equation that corresponds to the relation between causes and effects in the corresponding situation.” They were also asked to rate how familiar they were with each equation in any of its forms, on a scale of 1 to 7, where 1 meant that they *had never seen the equation* and 7 meant that they *knew the equation by heart*.

Results and Discussion

The results show that participants can reliably judge the causal form of equations. Table 1 shows the proportion of the participants choosing the most frequently chosen form of each equation. For the causal equations, the most frequently chosen form was always the expected causal form. For each participant, we calculated the proportion of times this modal version was chosen for the causal, the unclear, and the noncausal cases separately. The causal proportions ($M = .73$, $SE = .05$) differed from the other proportions [$M = .37$, $SE = .06$; $t(16) = 4.79$, $p < .001$] via a paired sample t test. We also compared the proportion of times the “none of them is causal” option was chosen for the noncausal equations versus the

causal and the unclear equations combined. The mean proportion for the noncausal equations ($M = .59$, $SE = .08$) was greater than the mean proportion for the other equations [$M = .2$, $SE = .04$; $t(16) = 5.59$, $p < .001$]. Finally, we tested the hypothesis that all choices, including the “none is causal” choice, should be chosen equally often. All of the causal and noncausal response distributions differed from uniform ($\chi^2s > 7.81$, $ps < .05$), whereas none of the unclear ones did.

All of the causal equations except for two had at least 65% of the participants choosing the expected causal form. In the case of *current = voltage/resistance*, the participants showed a bimodal distribution, with about half of them choosing current as the effect and the other half choosing voltage. Responses to *growth of crops = amount of rain/bags of pesticide* surprised us. Even though the causal form was the equational form chosen most often, most of the participants responded that none of the versions were causal. This equation did not follow this pattern in the other studies.

The causal equations were not the most familiar, so the strong preferences for a causal version cannot be attributed to familiarity. Also, as was expected, the unclear equations were the ones that people considered most unfamiliar. The mean familiarity ratings for unclear, causal, and noncausal equations were 1.67, 2.94, and 4.24, respectively.

STUDY 2

Study 1 suggests that people can reliably choose equational forms that match their causal models. If this is correct, the versions of equations that are solved for the variable that is considered to be the effect of the other two should be easier to grasp. Therefore, participants should choose the same versions as they did in the first study when asked which versions are most understandable. In the second study, we tested this hypothesis.

Method

Participants. Twenty undergraduates from Brown University came to the lab to answer this questionnaire, as well as some unrelated ones. None had participated in Study 1.

Procedure. Each participant received a questionnaire in which they were asked to choose the version of each equation they considered most understandable. This was defined as “the version they would teach to someone else because they find it most intuitive.” The rest of the questionnaire was similar to the one in the first study. It contained each version of each equation, as well as a “they are all equally understandable” option. The participants were also asked to rate how familiar they were with each equation, on a scale of 1 to 7.

Results and Discussion

As was predicted, Study 2’s understandability choice distributions showed the same patterns as the causal distributions in Study 1. Table 1 shows the proportion of participants choosing the modal form of each equation. The numbers in bold indicate that the modal form matched the form written in the table (the modal form for Study 1). The modal forms in this study matched the modal forms of the causal equations for all but two of the equations.

For each equation, we correlated the relative frequencies of selecting each form between Studies 1 and 2. The mean correlation across equations was .86 ($SE = .19$). As in Study 1, the pattern of understandability judgments cannot be attributed to familiarity, because the causal equations ($M = 2.97$) were less familiar than the noncausal equations ($M = 4.88$) and were not significantly different in familiarity to the unclear equations ($M = 2.06$). These results suggest a strong relation between causal interpretation and understandability of equations.

The data do suggest that factors other than causal structure contribute to judgments of understandability. More common forms are judged to be more understandable. This would explain why a noncausal form of Equation 8 was judged most understandable and why the canonical form of Equation 16 was chosen by almost all the participants in Study 2. We also observe that addition is considered more understandable than subtraction and multiplication more understandable than division. This could have contributed to the differences between Studies 1 and 2 in Equations 4, 8, and 9.

STUDY 3

The versions of equations considered causal tend to be the ones considered most understandable. If the hypothesis in this article is correct, this correlation is a result of the imposition of causal structure on the interpretation of mathematical forms. In the third study, we examined the relation between people's causal models and their ratings of causality and understandability by having participants produce causal graphs for each of the phenomena represented by an equation.

Method

Eighteen Brown and Yale University students answered the questionnaire. Only the 13 formulas in Table 1 with three variables were used in this study. Each participant received a questionnaire that asked, for each variable, which other variables would change as an effect of changing it. For example, for the equation $P = W/A$, they were shown the three variables, were asked for each one to "imag-

ine that someone changed that variable," and then were asked to circle the variables that they thought would change as a consequence. The participants in this study never saw the equations and were not told that these were variables from equations, so that they would answer the questions on the basis of their understanding of the phenomena, rather than using their knowledge of the mathematical forms.

Results and Discussion

We constructed aggregate causal graphs from the participants' responses. For each equation, any link that was chosen by more than half of the participants was included in the graph. The prediction was that all of the graphs from causally related variables would have two variables causing the remaining one.

Figure 1 shows the aggregate causal graphs that were generated for the causal equations. All three graphs show the predicted structure: the two predicted causes causing the predicted effect. In Graph Types A and B, when asked what would happen if the effect changed, the participants answered that neither of the other two variables would change. This violates mathematical constraints but follows from the causal models, suggesting that the participants were thinking about the phenomena, and not about the equations that represent them. One of the noncausal equations had a modal graph with no causal links, showing that it was perceived as noncausal. The other noncausal equation had two variables causing one, and that one causing the other two. The third column of Table 1 shows the causal graph for each equation. The fourth column of Table 1 shows the percentages of participants that had both causes causing the effect in their models.

We examined whether these graphs matched Studies 1 and 2's choice distributions. For each variable of each equation, we counted how many participants responded that the other two variables caused it. We then computed a correlation for each equation between the number of participants who had answered that a variable was an effect of the other two with the number of participants in each of the first two studies that had chosen the form of the equation that isolated that effect. The mean correlation with Study 1's causality distribution was .89 ($SE =$

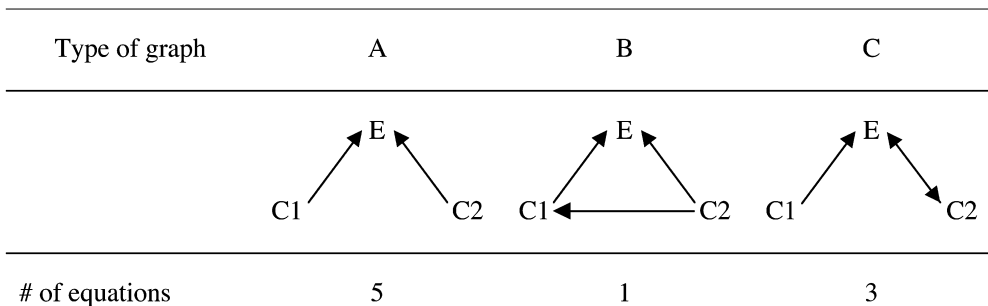


Figure 1. Causal graphs from Study 3. E stands for the predicted effect, C for a predicted cause. Causes point at effects. The number of causal equations that had each type as their aggregate causal graph are shown.

.25). The mean correlation with Study 2's understandability distribution was .80 ($SE = .23$). The correlations were high despite coming from a third sample of participants who had never seen the equations.

STUDY 4

The first three studies showed a strong relation among people's causal models, the versions of formulas they consider causal, and those they consider most understandable. All of these studies involved selection from a determinate set. In the fourth study, we tested whether people, when not given any cues about appropriate form, would write the causal version of an equation.

Method

Thirty-seven Brown University students in an introductory psychology class volunteered to answer the questionnaires. For this study, we used the three causal equations that were fabricated for earlier studies, so that answers could not be based on recall or familiarity (Equations 3, 4, and 6). These were the equations that were not taken from a textbook; rather, the variables were chosen because we expected people to have strong causal intuitions about how they relate. Predictions were based on the causal models generated in Study 3, which were consistent for these equations. For each equation, the participants were presented with the three variables in a counterbalanced order and were asked to write down the equation that they thought related them. For example, for Equation 6, the participants saw the following text, followed by a box with an equal sign in it:

A scientist has just derived the equation that relates the growth of crops (G), the amount of pesticide used (P) and the amount of rain during the week (R). Please write down in the box below the mathematical equation that you think relates these variables.

Results and Discussion

Table 2 shows that almost all of the participants wrote down equations that solved for the variable expected to be the effect. Each of the three distributions differed from uniform ($ps < .000001$). These results show that people naturally produce the causal version of an equation when given the freedom to produce any.

GENERAL DISCUSSION

The first study showed that people choose versions of equations that match the experimenters' causal models. The second showed that these versions are also the ones people consider most understandable. The third study showed that these results correlate highly with elicited causal models. Finally, the fourth study showed that peo-

ple have a strong preference for the causal version when asked to write down formulas. All the studies suggest that people find the versions of equations that match their causal models most intuitive. Their causal models appear to structure their preferred equational forms, presumably because they think of an equation as expressing a causal relation from a set of causes to a single effect, as predicted by the causal modeling framework. Factors other than causality also influence the understandability of equations. For example, entrenched forms are rated as more understandable (Experiment 2).

One potential alternative account of our results is that the participants preferred the equation that solved for the variable whose value they were most interested in, rather than for the effect per se. However, in the formulas used, no one variable seems to be any more important than the others, except insofar as it represents an effect rather than a cause. This account has the potential to explain the participants' selections of the most understandable form in Study 2, but it offers no leverage in explaining choices of the most causal form in Study 1 or the high correlations between distributions in Studies 1 and 2. Furthermore, if the critical variable were as subjective as interest or importance, one would expect more variability across participants than we saw. Our responses show a high degree of reliability. Finally, this account fails to explain why the forms selected in Studies 1 and 2 coincided so closely with the causal models obtained in Study 3.

Another possibility is that the participants chose equations that solved for the variables that were hardest to measure, because their values would be easier to obtain through calculation than through direct measurement. Again, the alternative fails. In some cases—for example, in Equation 2—the effect (the grade on an exam) is easier to quantify than its causes (understanding of the material and difficulty of the test), yet the participants predominantly chose the version that solved for the effect.

Would the results generalize to other causal model structures? For example, if two of the variables were effects of a third one, causal model theory would predict that people would represent the two links as two separate equations. This is a strong prediction that would have to overcome any preference for a single equation to be realized. We know of no tests of it as yet.

Not all formulas correspond to a causal model. Formulas sometimes serve other goals. For example, if the goal is to construct a scale, such as IQ, out of subscales, one might construct a regression formula that isolates IQ. Such regression formulas are designed to reflect a part-whole relation (the individual mental faculties that make up IQ are the parts and overall level of intelligence is the whole), rather than a causal relation.

Nevertheless, the findings in this article show that people sometimes impose causal structure when trying to understand noncausal information in the form of a mathematical formula. The value of causal models is fortified by the discovery that causality has a logic of its own (Pearl, 2000; Spirtes et al., 1993) and the observa-

Table 2
Mean Number of Participants Who Wrote Down Equations That Solved for the Predicted Effect (E) or One of the Causes (C1, C2)

	Variable Solved for		
	E	C1	C2
Mean number of participants	32	3	1.67

Note—Means of 37 participants across the three equations.

tion that people are capable of valid causal reasoning (Slovan & Lagnado, 2005). Knowing that learners are using causal knowledge should be an aid to those teaching mathematical reasoning in the sciences. Equations are useful not just because of their formal properties, but also because they reflect a reasoner's thinking process.

REFERENCES

- ANDERSSON, B. (1986). The experiential gestalt of causation: A common core to pupil's preconceptions in science. *European Journal of Science Education*, **8**, 155-171.
- 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.
- DISSA, A. A. (1993). Towards an epistemology of physics. *Cognition & Instruction*, **10**, 105-225.
- DRIVER, R., GUESNE, E., & TIBERGHIE, A. (1993). Some features of children's ideas and their implications for teaching. In R. Driver, E. Guesne, & A. Tiberghie (Eds.), *Children's ideas in science* (pp. 193-201). Buckingham, U.K.: Open University.
- GLYMOUR, C. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology*. Cambridge, MA: MIT Press.
- HALL, R., KIBLER, D., WENGER, E., & TRUXAW, C. (1989). Exploring the episodic structure of algebra story problem solving. *Cognition & Instruction*, **6**, 223-283.
- HINSLEY, D. A., HAYES, J. R., & SIMON, H. A. (1977). From words to equations: Meaning and representation in algebra word problems. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 89-105). Hillsdale, NJ: Erlbaum.
- HUNT, E., & MINTRELL, J. (1994). A cognitive approach to the teaching of physics. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 51-74). Cambridge, MA: MIT Press.
- LARKIN, J. (1983). The role of problem representation in physics. In D. Gentner & A. Stevens (Eds.), *Mental models* (pp. 75-98). Hillsdale, NJ: Erlbaum.
- NATHAN, M. J., KINTSCH, W., & YOUNG, E. (1992). A theory of algebra-word-problem comprehension and its implications for the design of learning environments. *Cognition & Instruction*, **9**, 329-389.
- PAIGE, J. M., & SIMON, H. A. (1966). Cognitive processes in solving word problems. In B. Kleinmuntz (Ed.), *Problem solving: Research, method, and theory* (pp. 51-119). New York: Wiley.
- PEARL, J. (2000). *Causality: Models, reasoning and inference*. New York: Cambridge University Press.
- PRIEST, A. G., & LINDSAY, R. O. (1992). New light on novice-expert differences in physics problem solving. *British Journal of Psychology*, **83**, 389-405.
- REIF, F., & ALLEN, S. (1992). Cognition for interpreting scientific concepts: A study of acceleration. *Cognition & Instruction*, **9**, 1-44.
- REINER, M., SLOTTA, J. D., CHI, M. T. H., & RESNICK, L. B. (2000). Naive physics reasoning: A commitment to substance-based conceptions. *Cognition & Instruction*, **18**, 1-34.
- SHERIN, B. L. (2001). How students understand physics equations. *Cognition & Instruction*, **19**, 479-541.
- SLOMAN, S. [A.], & LAGNADO, D. [A.] (2004). Causal invariance in reasoning and learning. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 287-325). San Diego: Academic Press.
- SLOMAN, S. A., & LAGNADO, D. A. (2005). Do we "do"? *Cognitive Science*, **29**, 5-39.
- SPIRITES, P., GLYMOUR, C., & SCHEINES, R. (1993). *Causation, prediction, and search*. New York: Springer-Verlag.
- WALDMANN, M. R. (1996). Knowledge-based causal induction. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), *The psychology of learning and motivation: Vol. 34. Causal learning* (pp. 47-88). San Diego: Academic Press.
- WHITE, B. Y. (1993). ThinkerTools: Causal models, conceptual change, and science education. *Cognition & Instruction*, **10**, 1-100.

(Manuscript received January 29, 2003;
revision accepted for publication December 18, 2003.)