Causal Models Frame Interpretation of Mathematical Equations

Daniel Mochon and Steven A. Sloman

Brown University

2431 words

Please address correspondence to:

Steven Sloman
Cognitive and Linguistic Sciences
Brown University, Box 1978
Providence, RI 02912
Email: Steven_Sloman@brown.edu
Phone: 401-863-7595
Fax: 401-863-2255

Abstract

We offer evidence that people 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 their expectations of change. When asked to write down equations, people have a strong preference for the version that matches their causal model.

Causal Models Frame Interpretation of Mathematical Equations

Empirical phenomena afford multiple representations that each highlight a different aspect of empirical structure. Even though mathematical equations can provide a concise and exact representation of how variables relate, they need not represent causal structure. Mathematical relations are symmetric, X = Y is identical to Y = X. Causal relations are not, X causes Y is not the same as Y causes X. Moreover, the representation of causal structure is essential to intervene on the world and influence its dynamics. In these senses, causal structure is more demanding than purely mathematical structure. Nevertheless, or perhaps as a result, people impose causal structure to understand (diSessa, 1993; Hunt & Minstrell, 1994; Reif & Allen, 1992). We propose they impose causal structure even when they need not: in the interpretation of mathematical representations. We hypothesize that people prefer equations whose variables are in an order that matches the causal relations that generate the phenomena represented.

People impose a causal frame even when it distorts a representation. For example, people assert a unidirectional causal relation between variables that interact in a more dynamic fashion. 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). Teaching correct causal interpretations often helps form a deeper understanding (White, 1993).

Pearl (2000) suggested that not all versions of equations are cognitively equivalent. People will learn and use equations more effectively to the extent that these match their causal model. Supporting evidence is that people consider the left side of an equation the outcome of something changing on the right (Sherin, 2001). The hypothesis predicts that, in the case of a three-variable equation matches a single causal model, the variable that is an effect of the other two variables should be written on the left and the two causes on the right. 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* $W = AP$. However, $P = W/A$ should seem more intuitive, because the most natural naïve causal model represents pressure as the effect of weight and area, not as a cause of them. 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.

Through a series of studies, we tested whether people have a preference for the version of equations that matches causal understanding. For these studies, we used 16 equations from various fields, such as physics, biology and economics (see Table 1 for full list of equations). They were grouped into 3 classes: The first type (Causal) were the equations with a clear causal model (e.g., $P = W/A$). These are the equations that represent causally related phenomena that are familiar, so most people understand what causes what. We assumed that the causal models of these simple situations are consistent across our population of Brown University undergraduates. The second type (Unclear) were equations with unclear causal models. These are equations that represent phenomena that are not very familiar, but for which causal models are assumed to exist. Therefore, we expected different people to think that they are related in different ways. One example is the equation that relates thermal efficiency (η), work done during one cycle (W), and heat added during one cycle (Q): $\eta = W/Q$. Some people might think that increasing the thermal efficiency would increase the work done during one cycle. Conversely, others might think that increasing the work done during one cycle would increase the thermal efficiency. The third type (Noncausal) are equations that have no causal model. These are equations that represent geometric relations or scale transformations with no causal interpretation. An example of this type would be the equation that gives the relation between temperature in Kelvin and Celsius: $K = {}^{o}C + 273$.

Study 1

The first study asked 18 participants to choose the versions of each equation that they considered causal. They were given the choice of all of the possible versions of each equation, as well as a "none of them are causal" option. They were also asked to rate how familiar they were with each equation. If equations are better understood to the extent they are related to people's causal models, people should be able to recognize the versions of the equations that match these models. Therefore, we predicted that the different types of equations should show different patterns. For the Causal equations, most participants should choose the version that we predicted to be the causal one. For the Unclear equations, there should be a uniform distribution across the different causal versions. Finally, most people should choose the "none of them are causal" option for the Noncausal equations.

The results were as predicted. The mean number of participants that chose "none of them are causal" for the Noncausal ($M = 10$, $S.E. = 1.53$) was significantly higher than the mean of the other equations ($M = 2.85$, $S.E. = 0.77$; $t(14) = 4.03$, $p < 0.01$). The mean number of participants that chose the predicted causal version for the Causal equations

(M = 11.44, S.E. = 1.25) was significantly higher than the mean for the modal choices of equations in the other conditions (M = 5.86, S.E. = 0.59; $t(11) = 4.04$, $p < 0.005$). Finally, all of the Causal and Noncausal response distributions differed from uniform ($p < 0.05$), while none of the Unclear did.

The Causal equations were not the most familiar ones, so the strong preferences for a causal version cannot be attributed to familiarity with certain versions of equations. Also, as expected, the Unclear equations were the ones that people considered most unfamiliar. On a scale from 1 (never having seen the equation) to 7 (knowing it by heart), the mean familiarity ratings for Unclear, Causal, and Noncausal were 1.67, 2.94, and 4.24, respectively.

Study 2

People understand equations to the extent that they match their causal models. In other words, the versions of equations that are solved for the variable they consider the effect of the other two should seem easier to grasp. Therefore they should choose the same versions as they did in the first study when asked which versions are most understandable. The second study tested this with a questionnaire that, except for the instructions, was identical to the one in the first study. In this questionnaire, 20 participants were asked to choose the version they considered most understandable, which was defined in the instructions as the version they would teach to someone else because they find it most intuitive.

As predicted, Study 2's understandability ratings showed the same patterns as the causal ratings in Study 1. A correlation between the relative frequencies of selecting each form of each of the equations between Studies 1 and 2 was computed. The mean correlation was 0.81. Again, the results cannot be attributed to familiarity because the Causal equations (M = 2.97) were less familiar than the Noncausal (M = 4.88), and not significantly different in familiarity to the Unclear (M = 2.06).

Study 3

The versions people consider causal correlate strongly with those they consider most understandable. If the hypothesis of this paper is correct, this correlation is a result of the imposition of causal structure on the interpretation of mathematical forms. The third study 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. The three equations that did not have three variables were not used in this study.

To elicit causal models, 19 participants were asked which other variables would change as an effect of changing the third variable. For example, for the equation $P = W/A$, they were shown the 3 variables and were asked for each one to imagine that someone changed that variable, and then were asked what would change as a consequence: one of the other variables, both, or none. 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 based on their understanding of the phenomena, rather than using their knowledge of the mathematical forms.

We constructed causal graphs from participants' responses. Every time, a participant answered that a variable (the effect) would change as a result of a change in another one (the cause), a causal link was drawn from the cause to the effect. We drew two types of causal graphs for each equation: the modal one across participants' individual graphs and one aggregating the responses across all participants. These two graphs coincided for 12 out of the 13 equations. The prediction is that the graphs would all have two variables causing the remaining one.

Figure 1 shows the causal graphs that were generated for more than one equation. All of the Causal equations had one of these three types. All three graphs show the predicted structure: the two predicted causes causing the predicted effect. Interestingly, in graphs types B and C, when asked what would happen if the effect would change, participants answered that neither of the other two would change. This violates mathematical constraints but follows from the causal models, suggesting that participants really were thinking about the phenomena and not the equations that represent them. One of the Noncausal equations had a modal graph with no causal links, showing that it was perceived as non-causal.

We tested the relation between these graphs and causality and understandability ratings by correlating the number of participants that answered that two variables caused the other one with the number of participants that chose the equation that was solved for that variable across equations. The mean correlation with Study 1's causality ratings was 0.90. The mean correlation with Study 2's understandability ratings was 0.79. The correlations were high even though they came from a third sample of participants who never saw the equations.

Study 4

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

For this study we used the three Causal equations that were fabricated for these studies so that answers could not be based on recall or familiarity. For each of the equations 37 participants were presented with the three variables, and were asked to write down the equation that they thought related them.

Table 2 shows that almost all of the participants wrote down equations that solved for the variable expected to be the effect. The three distributions each differed from uniform (p's < 0.000001). These results show that people naturally produce the causal version of equations when given the freedom to produce any.

Conclusion

The first study showed that people choose versions of equations that match the experimenters' causal model. 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 even when asked to write down equations, people have a strong preference for the causal version. All the studies suggest that people are most comfortable when thinking about the versions of equations that match their causal models. The value of causal models is fortified by the discovery that causality has a logic of its own (Pearl, 2000; Spirtes, Glymour & Scheines, 1993) and the observation that people are capable of valid causal reasoning (Sloman & Lagnado, 2002). Knowing that learners are using causal knowledge should be an aid to those teaching mathematical reasoning in the sciences.

References

Andersson, B. (1986). The experiential gestalt of causation: A common core to pupil's preconceptions in science. European Journal of Scientific Education, 8, 155-171.

DiSessa, A.A. (1993). Towards an epistemology of physics. Cognition and Instruction, 10, 105-225.

Driver, R., Guesne, E., & Tiberghien, A. (1993). Some features of children's ideas and their implications for teaching. In R. Driver, E. Guesne & A. Tiberghien (Eds.), Children's ideas in science. (pp. 193-201). Buckingham, United Kingdom: Open University.

Hunt, E., & Minstrell, 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: M.I.T.

Pearl, J. (2000). Causality: Models, reasoning and inference. New York: Cambridge University Press.

Reif, F. & Allen, S. (1992). Cognition for interpreting scientific concepts: A study of acceleration. Cognition and Instruction, 9, 1-44.

Reiner, M., Slotta, J.D., Chi, M.T.H., & Resnick, L.B. (2000). "Naïve physics reasoning" a commitment to substance-based conceptions. Cognition and Instruction, 18, 1-34.

Sherin, B.L. (2001). How students understand physics equations. Cognition and Instruction, 19, 479-541.

Sloman, S.A., & Lagnado, D. (2002). Counterfactual undoing in deterministic causal reasoning. Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society, MD, USA.

Spirtes, P., Glymour, C. & Scheines, R. (1993). Causation, prediction, and search. New York: Springer-Verlag.

White, B.Y. (1993). Thinkertools: Causal models, conceptual change, and science education. Cognition and Instruction, 10, 1-100.

Authors' Note

Daniel Mochon and Steven A. Sloman, Brown University. We thank Dave Lagnado and Dave Sobel for their help on this project. This work was funded by NASA grant NCC2-1217.

Table 1

One form of each equation. For Causal and Unclear types, causal form shown.

Causal Equations

1. behavior potential= expectancy of the reinforcement *value of reinforcement
2. grade = understanding/difficulty of the test
3. growth rate = birth rate – death rate
4. neuron firing rate = amount of excitatory enzyme/ amount of inhibitory enzyme
5. price = money demand/money supply
6. growth of crops = amount of rain/bags of pesticide
7. force = mass/acceleration
8. voltage = current*resistance
9. pressure = weight/area

Unclear Equations

10. wavelength = Planck's constant/mass*speed
11. thermal efficiency = work done during one cycle/heat added during one cycle
12. heat capacity = change in heat/change in temperature
13. axial deformation = internal column load*length/cross section area*Young's modulus

Noncausal Equations

14. z = x/y
15. Kelvin = Celsius degrees + 273
16. length of the hypotenuse^2 = length of one side^2 + length of the other side^2

Note. The equations were not presented in this format to the participants. They were presented as 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 of the other equations are actual equations from their respective fields.

Table 2

Mean number of participants who wrote down equations that were solved for the predicted effect (E), or one of the causes (C1, C2).

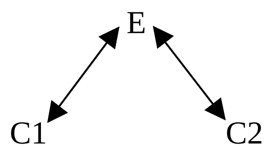
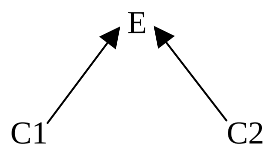
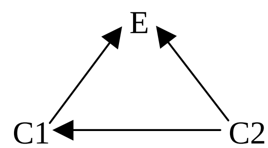
Variable solved for	E	C1	C2
Mean number of participants	32	3	1.67

Note. Means across the three equations of 37 participants.

Figure 1 Caption

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

Figure 1

Type of graph	A	B	C
			
# of equations	7	2	2