

Stephens, A. L., Roderick, S., Shin, N., & Damelin, D. (2023). Students do not always mean what we think they mean: A questioning strategy to elicit the reasoning behind unexpected causal patterns in student system models. *International Journal of Science and Mathematics Education 21*, 1591-1614. <https://link.springer.com/article/10.1007/s10763-022-10308-z>

Students don't always mean what we think they mean: A questioning strategy to elicit the reasoning behind unexpected causal patterns in student system models

ABSTRACT: An ability to engage in systems thinking is necessary to understand complex problems. While many pre-college students use system modeling tools, there is limited evidence of student reasoning about causal relationships that interact in diverging and converging chains, and how these affect system behavior. A chemistry unit on gas phenomena was implemented in two successive years with 73 high school students. Although the phenomena could be explained with simple linear causal reasoning, many student models included surprising and problematic causal chains and non-linear patterns. Commonly, discussion about student models in classrooms and interviews focuses on individual causal relationships. However, in our experience, this can fail to bring to the surface conceptual issues that result from the combination of two or more relationships. Using interview data from 19 students, we looked for instances where students explained their conceptions and we identified questioning strategies that elicited these explanations. Questions that appeared most productive asked about distal relationships between nonadjacent variables in model substructures of three to four variables. In response to these questions, eight students expressed their thinking about relationship combinations, including unexpected reasoning that had not emerged during classroom instruction. Using exemplars from three interviews, we argue that students engage in complex causal reasoning that may be implicit and unexpected, and that if this is not recognized, it cannot be responded to during instruction. We suggest that using model substructures as a mutual visual referent along with a simple questioning strategy shows promise for helping students make their causal reasoning explicit.

KEY WORDS: causal patterns, questioning strategy, system modeling, systems thinking, eliciting student ideas

Many crucial issues today involve complex systems. Although systems thinking can help citizens take effective action in these areas, studies suggest that even highly educated adults generally have poor understanding of basic systems concepts (e.g., Booth Sweeney & Sterman, 2000). System models have been proposed as useful tools to help students understand systems and convey that understanding to others (National Research Council, 2012). Semi-quantitative system modeling tools have been developed to provide an onramp to system modeling that does not require the use of equations (e.g., Metcalf et al., 2000). However, there is not yet a robust pedagogy for teaching system modeling to pre-college students, particularly with semi-quantitative tools. Here we explore how such modeling tools can be used to surface students' causal reasoning.

Although much work has been done on detecting student reasoning with the use of conceptual models (Ben-Zvi Assaraf & Orion, 2005), concept maps (Tripto et al., 2016), and agent-based models (Stroup & Wilensky, 2014), with the exception of Hmelo-Silver et al.'s (2017) components, mechanisms, and phenomena (CMP) representation, these have had limited success in eliciting student reasoning about *combinations* of relationships. While there is evidence that teachers using a semi-quantitative system modeling tool with their students shifted toward asking more questions about causal links and mechanisms compared to when they used paper modeling (Nguyen & Santagata, 2021), in our experience the adoption of a semi-quantitative modeling tool does not guarantee that teachers will move the focus of their questions beyond individual model elements. We have found that questions about individual relationships are not always sufficient to reveal the reasoning behind students' modeling choices, and the models themselves may not reveal what students are thinking.

EXAMPLE: Derek and his partner created a model that included two interacting causal pathways, one predicting that raising the temperature of air inside a fixed volume

would cause air particle speed to increase and the other predicting that raising the temperature would cause air particle speed to decrease. Investigators surmised that the students either did not understand some of the modeling conventions or they had forgotten that the volume was fixed. In an interview, when probed about the individual relationships, Derek provided plausible justifications for each one. Trying different probes, the interviewers discovered he also had a plausible justification for the combination of relationships in his model, rooted in a non-canonical conception about what produces pressure. This conception had not surfaced during the lesson.

Our objective in this study is to identify probes that can help elicit the student thinking that underlies complex causal structures in their models. Using interview data from three students who participated in a chemistry unit on gas phenomena to exemplify findings from a larger set of interviews, we will suggest that identifying and describing a questioning strategy that can move the focus of discussion beyond individual causal relationships could be of use to teachers as well as to researchers. Although the interview context does not always translate to the classroom, similar skills are needed to detect student thinking (Lee, 2021). Transcript excerpts, along with student models, help us identify the reasoning behind relationship combinations in the models and describe the kinds of questions that were successful in eliciting this reasoning. We will also suggest that pairing this questioning strategy with visual affordances in the semi-quantitative system models can help students articulate rich, inchoate, and sometimes surprising aspects of their systems thinking.

Relevant Prior Work on Eliciting Systems Thinking

Challenges of Teaching and Learning Systems Thinking

The Next Generation Science Standards (NGSS), widely used in the United States, identifies *Systems and System Models* as a crosscutting concept to be introduced in grades K-2 and recommends that students begin developing their own system models by grades 9-12, if

not before (NGSS Lead States, 2013). Several issues complicate the attainment of these NGSS goals. Most learners are familiar with only a narrow range of causal patterns, such as A affects B affects C (Perkins & Grotzer, 2005). Studies in scientific reasoning have focused on the ability to identify which variables are producing an effect, but have not asked whether, having done so, an individual can reason about the simultaneous effects of multiple variables (Kuhn, 2007). When asked to predict the effects of new combinations of variables, preadolescents, as well as some adults, perform poorly (Kuhn & Dean, 2004). Beyond this, complex systems thinking and modeling require a wide array of cognitive resources and abilities (e.g., Ben-Zvi Assaraf & Orion, 2005), including knowledge about the nature of models, domain knowledge, and scientific reasoning skills (Hmelo-Silver & Azevedo, 2006).

Supporting modeling can be challenging (e.g., Hokayem & Schwarz, 2014), and science teacher education typically has not focused on modeling (system or otherwise) (Gilbert, 2004). Teachers typically have a weak understanding of complex systems concepts (Yoon et al., 2018). Fisher (2018) is concerned that teacher training is siloed so that teachers are not prepared for the broader background needed to facilitate discussions that support system modeling. Educators and learning scientists need to better understand which aspects of learning about complex systems need scaffolding as well as how to address those needs (Hmelo-Silver & Azevedo, 2006).

Advantages and Challenges of Semi-Quantitative System Models

In the 1990s, several modeling tools attempted to simplify the learning of complex systems by using variables linked by semi-quantitative causal relationships. These include *IQON* (Bliss, 1994; Miller et al., 1993), *Model-It* (Metcalf et al., 2000; Stratford et al., 1998), *LinkIt* (Ogborn, 1999), and, more recently, *ModelsCreator* (Komis et al., 2007). Stratford et al. (1998) found that secondary students who used *Model-It* engaged in a wide range of cognitive strategies and most engaged in causal and correlational reasoning as they created

connections and defined relationships between variables (“factors”). However, Bliss (1994) noted that secondary students had trouble seeing that if the value of one variable was changed, it could affect the values of many others.

As system models become more complex they can contain causes that produce multiple effects and/or effects that become causes, as in branching patterns and causal chains. These interaction patterns are described in the literature (Pearl & Mackenzie, 2018; Perkins & Grotzer, 2005), but when we saw them in student models it was not clear that the students always used them appropriately. Students find ways to ignore or reduce complexity in such situations (Rozier & Viennot, 1990). For instance, undergraduates studying thermodynamics tend to neglect to consider which variables are being held constant (Kautz et al., 2005).

Importance of Eliciting Student Ideas

Awareness of the importance of eliciting student ideas is not new (e.g., Hammer, 1996; van Zee & Minstrell, 1997). According to Hogan and Thomas (2001), “Educators need to be able to diagnose students’ thinking while modeling, and intervene in appropriate ways at opportune times to support and enhance their learning” (p. 344). However, even when student ideas have been elicited, teachers often miss opportunities to ask follow-up questions (e.g., Pimentel & McNeill, 2013; Weiland et al., 2014). Discourse strategies developed to support scientific reasoning include initiation-response-feedback (IRF) (Lemke, 1990; Nassaji & Wells, 2000), argumentation (Osborne et al., 2004), and focusing on causal mechanisms (Russ et al., 2009). In our experience, use of these strategies, though valuable, does not help teachers move the focus beyond individual causes to more complex ones.

One way to elicit student ideas is to help them make their thinking visible (Collins et al., 1991). Although having students construct models should help, our own work (Stephens & Ki, 2017) has shown that students’ written explanations of their models do not always agree with what their models seem to indicate. If there are layers of thinking underlying

student system models that we are missing, we suggest it is critical to figure out how to detect and understand them so that we can develop ways to support students more effectively.

Questions and Other Prompts to Elicit Thinking About Multiple Relationships

Because the use of explicit prompts can elicit student explanations and help them reflect on their own thinking (Chi et al., 1994; Tripto, et al., 2016), we looked for questions and other prompts that have successfully elicited student reasoning about interactions among causal relationships of the kind we frequently see in student system models (such as interacting causal chains). Our initial review of the literature did not uncover these kinds of prompts in the context of system modeling. We expanded our search to include any kinds of prompts designed to elicit student thinking about system models, including written assessment questions. We found many studies that reported prompts to elicit aspects of students' system thinking used in conjunction with visual representations of systems (Gilissen et al., 2020; Hmelo-Silver et al., 2017; Levy & Wilensky, 2004; Stroup & Wilensky, 2014; Wilensky, 2003). Although some included hierarchy and feedback, we did not find reports of instruments or prompts that attempted to elicit student reasoning about interactions among multiple causal relationships of the kind we were seeing. For example, Gilissen et al. (2020) asked secondary students in interviews or writing to identify seven systems characteristics in an image of a system after enactments of lessons using a physical model. These included aspects of hierarchy of a system and feedback but did not directly address interacting causal relationships.

The most extensive literature we found about prompts with modeling concerned the use of concept maps to elicit and assess system thinking. Concept maps have node-link assemblies with links represented by labeled arrows, though these links need not be causal. However, for the most part, in the literature we reviewed, these assessments did not focus on interactions among multiple causal relationships. Yin et al. (2005) supplied a list of (non-

causal) linking phrases to subjects. Hrin et al. (2017) used partially filled diagrams that pre-established the structure. Brandstädter et al. (2012) provided subjects either concepts and linking words or a verbal description of a system and evaluated individual relationships in the concept maps students constructed from them. Although Sommer and Lücken (2010) assessed combinations of relationships using concept maps and pre-post questions with elementary school children, probes into student thinking about interacting causal chains were not reported. In a study by Plate (2010) using causal maps, node-link assemblies where all relationships were semi-quantitative and causal, students in 7th and 8th grade identified individual relationships between pairs of concept cards in a stack of 36 cards. Analysis of configurations that emerged as a result of these individual relationships identified linear, simple branching, closed branching, and causal loops and whether a connection between a given pair of concepts was a direct connection or in the form of a causal chain. Notably, the students themselves did not identify these configurations and this analysis did not focus on the thinking behind those configurations.

A few studies used concept maps along with interviews to elicit and assess system thinking and showed some success in eliciting causal chains and/or causal loops. Ben-Zvi Assaraf and Orion (2005) used a battery of 10 instruments including concept maps, drawings, Likert-style questionnaires, and interviews to assess system thinking of 8th grade students about the hydrocycle. Student ability to create a framework of relationships was assessed from pre- and post-concept maps, and their ability to think in terms of a cycle with a Likert questionnaire of statements the students agreed or disagreed with. In interviews, students were asked to elaborate on their drawings and answers to the written instruments and completed additional tasks. Specific questions used in the interviews were not reported in this study, but in a similar study (Ben-Zvi Assaraf & Orion, 2010), the same authors were able to elicit descriptions of causal chains about the hydrocycle from some 4th grade students by

asking which of the elements in their drawings and repertory grids influenced other elements and which they were influenced by. Tripto et al. (2016) conducted in-class reflection interviews with 11th grade students about concept maps they had created the previous year. One purpose was to see whether students could organize relationships they had identified into a coherent framework of relationships. Explicit questions about interactions worked better than implicit questions, but still resulted in only 17% of students describing phenomena as arising from multiple interactions.

We searched the literature for written and spoken questions and other prompts that could elicit the reasoning behind the kinds of complex causal structures we were seeing in students' semi-quantitative system models. For the most part, the studies above either did not focus on eliciting these kinds of reasoning or had limited success in their attempts to do so. Our goal is to identify a questioning strategy not only to assess this reasoning once it is fully formed, but to elicit thinking that may still be implicit. Such a strategy should suggest a useful way for teachers and researchers to understand whether hidden causal thinking may be affecting students' progress toward engaging in complex systems thinking and modeling.

Research Questions

How can we elicit the thinking underlying causal patterns in students' system models? In particular, how can we scaffold articulation of student thinking that goes beyond consideration of individual relationships in a system but may be implicit or inchoate?

Methods

Context and Curriculum

The focus of Building Models (a National Science Foundation [NSF] design-based research and development project) was to explore how students learn to engage in systems thinking and system modeling. The project developed SageModeler, a new semi-quantitative modeling tool designed to make system modeling more accessible for younger students, and

supported teachers to enact co-developed science units that included system modeling.

Although the issues raised and the questioning strategy we identify are not content specific, an overview of the content will help contextualize the student utterances during interviews.

Before beginning a unit exploring a phenomenon related to properties of gasses, students completed a five-day introductory unit about systems and modeling using the new modeling tool. At the beginning of the gas phenomena unit, which included 11 lessons of approximately 70 minutes each, students viewed a video of an oil tanker imploding. They learned about factors that could have contributed to the phenomenon through hands-on experiments, computer simulations, readings, and class discussions. When creating their system models, students were asked to think of factors related to the properties of gasses that could be measured on a low-to-high scale and that could affect the chances for implosion, including changes in temperature, changes in amount of gas inside the tanker (due to steam condensing), consequent changes in gas pressure, and the fact that the atmospheric pressure on the tanker did not change appreciably. Students were encouraged to consider gas pressure as an emergent property of molecular collisions. They also learned that models cannot include all aspects of a phenomenon, so they must choose which critical factors to include. Students revised their models throughout the unit as they learned new aspects of the phenomena or engaged in experiments. Each of the revisions was saved and students wrote explanations for any changes they had made since the previous revision. Although no specific target model was provided, it was expected that in some way students would represent a mechanism that would explain how lowering the temperature inside the sealed tanker would increase the chance that it would implode (Figures 1 and 2).

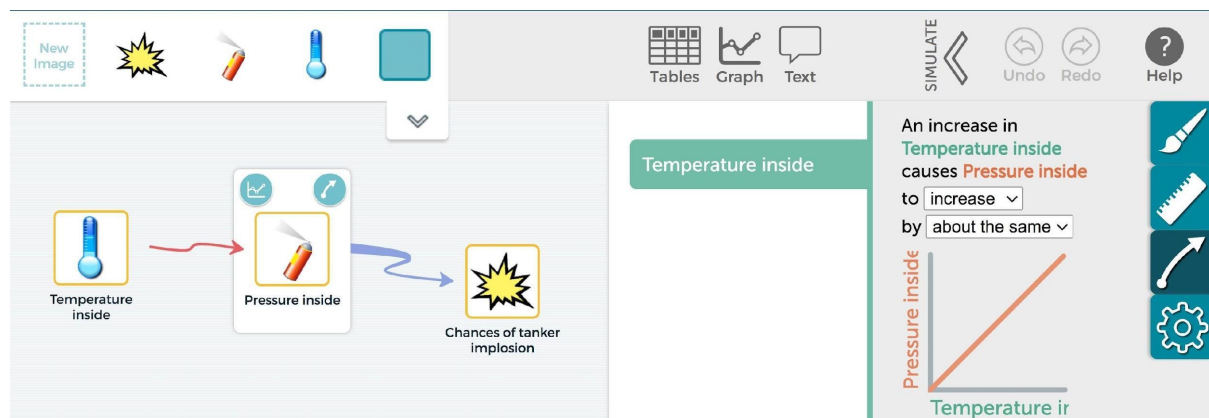
Enactments of this unit were chosen for study because they provided a rich source of data on student thinking about the complex causal patterns we observed in models across multiple units in the project (Stephens & Roderick, 2019).

The Modeling Tool

The models were constructed using SageModeler, a semi-quantitative web-based modeling tool. As shown in Figure 1, each arrow represents the effect of one variable on another, with red arrows indicating a change in the same direction, a positive effect; if the *Temperature inside* increases, so will the *Pressure inside*. A blue arrow indicates a change in the opposite direction, a negative effect; as the *Pressure inside* increases, the *Chances of tanker implosion* go down. These cause and effect relationships are represented in multiple forms: as arrows, words, and graphs. In Figure 1, the first drop-down menu in the relationship box (“increase,” “decrease,” “vary”) determines the color of the arrow, while the second (“about the same,” “less and less,” etc.) determines the shape of the arrow tail.

Figure 1

Variables and Relationships in a SageModeler Model



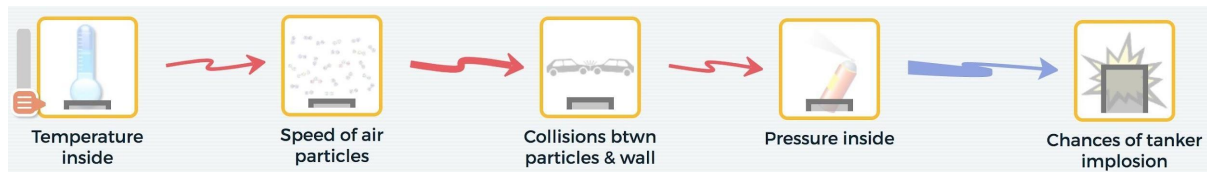
Each relationship between variables is represented in the model by an arrow, and in the relationship box (right) by a sentence and an automatically generated graph.

Clicking “Simulate” causes a slider to appear on any independent variable, as in Figure 2. Students can use the sliders to examine how modifying inputs can impact other variables in the model, with model state represented by bar graphs within the variable icons. In Figure 2, the slider for *Temperature inside* has been moved to the bottom of its range and the effect on *Chances of tanker implosion* can be seen. This model predicts that as the

Temperature inside decreases, the *Chances of tanker implosion* will increase. Students can also create graphs that show relationships between variables that are distal¹ in the models.

Figure 2

Possible Structure for a Final Model



It was expected that in some way students would represent that lowering the temperature inside the sealed tanker would increase the chances that the tanker would implode.

Participants and Data Sources

The research site was in a suburban high school in the northeastern U.S. chosen based on its proximity to the research institution. A chemistry teacher ran the gas phenomena unit in one section the first year and four sections the second year. Seventy-three 10th grade students participated. In each enactment we asked the teacher to select focus students who would be the subject of additional data collection including interviews, part of which would be dedicated to understanding student thinking around their models. These students were selected by the teacher to represent a variety of abilities and performance levels to provide a sample as representative of the classes as possible. In Year 1 there were 6 focus students and in Year 2 there were 16 (4 per section), for a total of 22 focus students. We successfully scheduled interviews for 19 of these students. The students had worked in pairs but all were interviewed individually. Post-interviews were conducted the day after the unit ended.

To prepare for the interviews, the first and second authors reviewed the models created by those students along with their written explanations for each saved model revision. The semi-structured interviews were conducted by the first author alone or with the third author, both of whom had participated in class sessions helping to implement the lessons. The interviews began with a structured segment in which the student was asked what they

recalled about the phenomenon and its causes, and what they remembered about their most recent model. The student then displayed that model on a computer screen and an interviewer asked them to “talk me through your model.” If needed, the student was asked to clarify or expand their description. Once the interviewers were satisfied they understood the student description, the student was asked an individualized question concerning something in the model the interviewers found puzzling or that appeared inconsistent with the student’s written explanations. Finally, the interviews moved to probing questions as they sought to more deeply understand the thinking that had gone into the production of the model. The interviews closed with questions about the students’ experiences with the unit.

In a prior study (Stephens & Roderick, 2019) we analyzed student models from both years for their structural characteristics. We identified 11 causal patterns, many of them complex, but also found that the complexity of the models was not always associated with the system being modeled. We frequently observed this teacher helping students evaluate the validity of individual links but less often considering larger causal structures. The current study investigates the thinking underlying the students’ creation of the four complex causal structures identified as the most prevalent in student models in the prior study. (See Table 1.)

Data Analysis

We began by reviewing all 19 of the interview transcripts. For data reduction, we used purposeful sampling, a technique to identify interviews for which recorded discourse could “purposefully inform an understanding of the research problem and central phenomenon in the study” (Creswell, 2007, p. 125). For instance, if the models of a student pair did not contain any of the patterns described in Table 1, or if a student had been absent and did not understand the final model their partner had produced, their interviews were excluded. Eight interviews were judged suitable for further analysis. To analyze these interviews, we first identified all areas in the transcripts where students engaged in reasoning

beyond explaining individual relationships (e.g., the combined effect of two relationships). We then identified the interviewer questions that preceded this reasoning. For reasons of parsimony, we use three subjects to exemplify the results. These three were chosen because between them, their interviews provide examples of all four patterns (Table 1) and represent the variety of reasoning we observed across the set of eight interviews. Transcript excerpts illustrate how student articulation of broader system-level reasoning may be supported even when this reasoning is nascent. In two instances some interviewer questions were followed by reasoning about individual features only. We contrast these with questions in the same interviews that were followed by reasoning about combined effects. Verification procedures included the use of multiple investigators, interviews from two different years, discrepant cases, and discussions of tentative interpretations with colleagues.

The Causal Patterns

Table 1

Four Patterns in Student Models

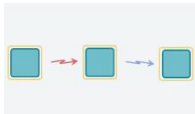
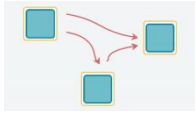
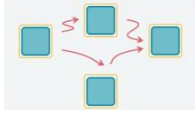
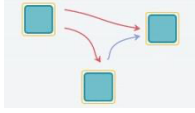
Pattern	Representation	Description
A. Causal chain		Chain of two or more cause and effect relationships
B. Single link paired with causal chain		Two variables connected by both single link and causal chain
C. Parallel causal chains		Two variables connected by two causal chains
D. Compensating effects		Effect of one variable on another variable is both positive and negative

Table 1 shows the causal patterns most prevalent in the model sequences examined in Stephens and Roderick (2019). We focus here on instances where these patterns did not make

sense in the context of the models. Pattern A is a causal chain of at least three variables. Most of the final gas phenomena models analyzed contained at least one chain where the effect of the first variable on the last variable in the chain did not match the scientific description of the phenomenon. All occurrences of patterns B, C, and D in the gas phenomena models were judged problematic. Pattern B occurred in about half of the final models, usually with one of the links unneeded, and Pattern C occurred in about a third of them, often with one of the chains unneeded. Pattern D, occurring in about a sixth of the final models, indicated that an increase in one variable would cause another to both increase and decrease. These patterns were not a focus of instruction; they had not been identified as problematic at that time.

Results: Questions About Complex Causal Patterns

During analysis of the interview data, we anticipated being able to identify a specific kind of prompting question for each causal pattern. However, we found that the questions preceding student reasoning behind the patterns seemed to arise from a single strategy, that of focusing on combined effects of *distal* relationships, such as relationships between non-adjacent variables in a causal chain. From the interviews of Mark and Derek the first year, and May the second year (all pseudonyms), we describe several examples of student reasoning and the interviewer questions that preceded the reasoning. In the last example, we will revisit the episode described in the introduction, about Derek's contrasting predictions.

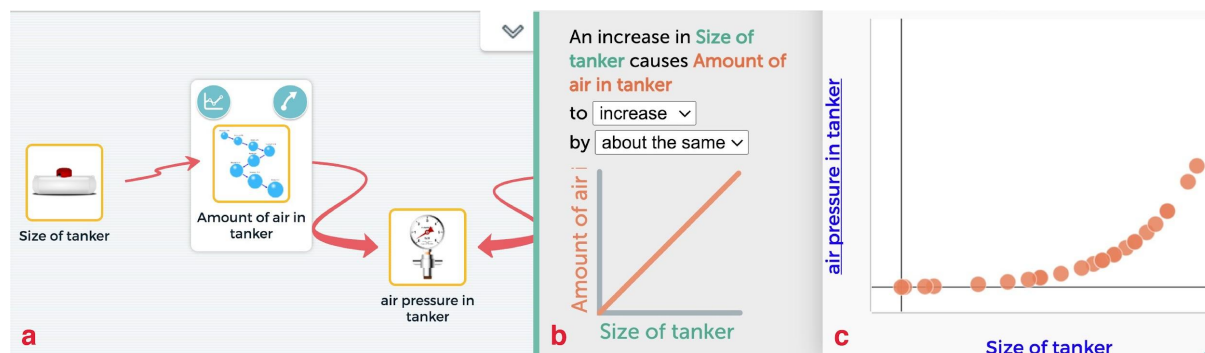
Mark: Problematic Causal Chain (Pattern A)

The most common problematic pattern in the models was variables arranged in chains of relationships (mostly causal) where the resulting effect of the first variable on the last one did not agree with the canonical scientific explanation. Although students tend to reason in sequential chains as opposed to reasoning in more complex interaction patterns (Perkins & Grotzer, 2005), we anticipated that students might have trouble reasoning with long chains.

However, at times we found that even two-link chains (three variables linked in series) posed challenges for these high school students.

Figure 3

Problematic Causal Chain in Mark's Final Model



- a) Two-link causal chain in the model. b) Relationship panel with representations of the single link between Size of tanker and Amount of air in tanker. The graph is auto-generated. c) A student-generated graph of the effect of Size of tanker on air pressure in tanker.

Figure 3a is a two-link chain in the final model produced by Mark and his partner. They chose *air pressure in tanker* as the outcome (the final effect) of their model rather than explicitly including a variable for chances of collapse. Two causal chains led to this outcome variable. The chain to the right concerned the effects of temperature and was as expected except for one unneeded variable. It is not shown for considerations of space. The short chain to the left of this variable was problematic and is shown in full. The combined effect of the two links is that increasing the size of the tanker would increase the air pressure in the tanker. This can be considered a prediction made by their model, though a surprising one. In written explanations of their modeling decisions, neither Mark nor his partner referred to this result. In the final interview, when asked to talk through his model, Mark initially described the two links separately, using the non-specific phrase “connects to” to refer to the relationships, “Over here on the left side we have *Size of the tanker* and that connects to the *Amount of air in the tanker*, and *Amount of air in the tanker* connects to the *pressure inside of the tanker*.”

Later, the interviewer (“I”) asked a clarifying question about the two-link chain.

I: “What I’m asking is, if you trace the cause and effect between *Size of tanker* and what effect that has on *Amount of air* and what effect that has on *Air pressure*, according to your model—”

Mark: “Oh yeah.”

I: “—if you change the *Size of tanker*, what would happen to the *Air pressure*?”

At this point Mark brought up the relationship box representing the first link (Figure 3b). As he talked, he moved his cursor over the graph in this box.

Mark: “So this arrow shows a linear relationship, so this is saying that the bigger the size of the tanker, the larger the amount of air in the tanker. So if there’s more air in the tanker, I’m guessing that it would have a higher air pressure. So, that’s what that arrow is showing. So, yeah, yeah. It says that if the tanker is larger, there’s more air. Which would affect the pressure.”

This time Mark described the links in causal terms, but in the last sentence he substituted the neutral verb “affect” rather than specifying the direction of effect. His hesitation and qualifications suggested that he may not have previously reasoned about the combined effect of the two relationships represented by these links. Although the interviewer asked about the combined effect, at this point she did so in a step-by-step manner. That did not appear to be sufficient to help Mark reason clearly about this part of his model. Later the interviewer asked Mark about a graph he had constructed with his partner using model output to plot *Size of tanker* directly against *air pressure in tanker* (Figure 3c).

I: “So what does that graph predict about the relationship between *air pressure in tanker* and *Size of tanker*?”

Mark: “This says that as the size of the tanker gets higher, the air pressure gets higher, which is wrong! That’s wrong.”

I: “Ah!”

Mark: “It also depends on if the tanker is open or closed. Because if it were closed, then we’d have the same amount of air molecules, and in that case, the air pressure would actually be lower as the size of the tanker got higher.”

We can infer that when reasoning directly about the relationship between the two non-adjacent variables, the student realized not only that his model was predicting something

contrary to other information he had about the system, but also that the error involved a quantity (*Amount of air in tanker*) that should have been held constant when considering that relationship (the effect of *Size of tanker* on *air pressure in tanker*).

Student Issue and Questioning Strategy That Preceded the Response

Mark exhibited challenges with causal reasoning when two relationships were combined in a causal chain, which is consistent with thinking identified by Kautz et al. (2005) as problematic for undergraduates in science courses such as thermal physics. Because students tend to lose track of variables being held constant, they would misapply gas laws, interpreting relationships between pairs of variables as always holding regardless of changes in other variables. Similarly, Mark appeared to lose track of constants: *Amount of air* in the volume/pressure relationship and *Size of tanker* in the *Amount of air/air pressure* relationship.

When the interviewer asked Mark to trace cause-and-effect along a two-link causal chain while looking at that part of his model, Mark did not notice the discrepancy but reasoned separately about each individual relationship. However, when the interviewer drew Mark's attention to a graph the student had already created and asked about the distal relationship represented there, Mark realized his model predicted that an increase in the size would cause an increase in the pressure. This appeared to serve as a clue for him to take another look at this part of his model. We suggest that drawing Mark's attention to a distal relationship helped him articulate and clarify his thinking about that aspect of the phenomenon and how he had represented it.

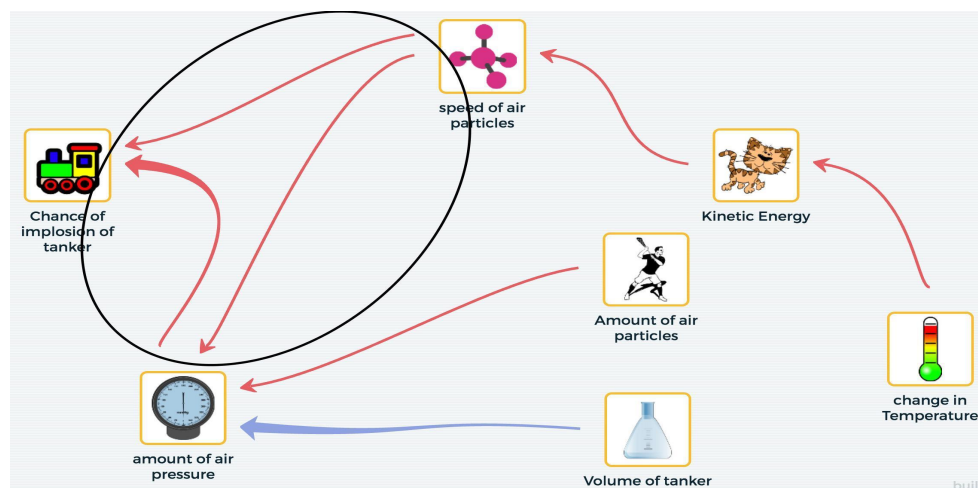
Derek: Single Link Paired with a Causal Chain (Pattern B)

As students learn more about a phenomenon, they often add new variables to their models. However, we frequently observe students adding new variables and links while retaining all existing links. This often results in two causal pathways between a pair of

variables, one of which is the old single link, a *proximal* link, and the other a new causal chain, a distal link between the same two variables (Figure 4). This was the second most common problematic pattern in these models.

Figure 4

Single Link Paired with a Two-Link Chain in Final Model of Derek and His Partner



There are two causal pathways from speed of air particles to Chance of implosion of tanker.

Derek's final model contained this pattern (circled in Figure 4). In his final interview, he was asked about it by two interviewers ("I1" and "I2").

I1: "I have a question about *speed of air particles*. It looks as though *speed of air particles* directly affects *Chance of implosion of tanker* and then also affects the *amount of air pressure*, which affects the *Chance of implosion*. Are you saying that the *speed of air particles* has more than one effect?"

Derek: "Yeah, but now looking at it, I probably could delete one of those arrows."

I2: "Which one would you delete?"

Derek: "I'd probably delete the direct relationship between *Chance of implosion of tanker*. (pause) 'Cause I feel like there's something in between *speed of air particles* and *Chance of implosion*."

The link between *speed of air particles* and *Chance of implosion of tanker* did not represent any additional information not already represented in the model and the student decided that it could be removed. (The problematic prediction about the effect of *amount of air pressure* on *Chance of implosion of tanker* was addressed later in the interview. Other aspects of this model will be discussed in a later section.)

Student Issue and Questioning Strategy That Preceded the Response

Derek and his partner were still becoming accustomed to modeling conventions. They had added new variables and links to their model and not noticed that the link in question was now unnecessary.

The interviewer asked a question of the form “Does Variable A have more than one effect on Variable B?” This can draw the focus away from individual relationships toward a specific combination of relationships. Once Derek’s attention was drawn to these, he began to reevaluate that part of his model.

May: Single Link Paired with Causal Chain, Compensating Effects (Patterns B and D)

If a pathway between two variables indicates the first variable will have a positive effect on the second, while a different pathway between the same two variables indicates the first will have a negative effect on the second, the effects will tend to cancel when the model is simulated (Figure 5). However, it may be unclear that this is what the student intended. This pattern was not as common as the others, but did appear in several final models.

In the previous two examples, the students came to a better understanding of the model and/or of the science when responding to the interviewer questions. In May’s interview it is doubtful she gained in her understanding of either. However, the questions elicited articulation of deep issues she had with conceptualizing the phenomenon, issues that had not been evident in the classroom.

The interviewer began by focusing on the two pathways between *Weather condition* and *Temp Inside the tanker* because it was not clear what idea was being expressed, asking first about the three individual links (Figure 5). These questions elicited separate descriptions of the conditions inside and outside the tanker (hot steam inside and cold outside), but did not make clear May’s thinking about the two causal pathways. She explained that the blue arrow meant the cold air outside would cause the inside to cool. The interviewer probed May’s

understanding of this relationship. (In the following, ellipses denote repetitive utterances omitted for clarity.)

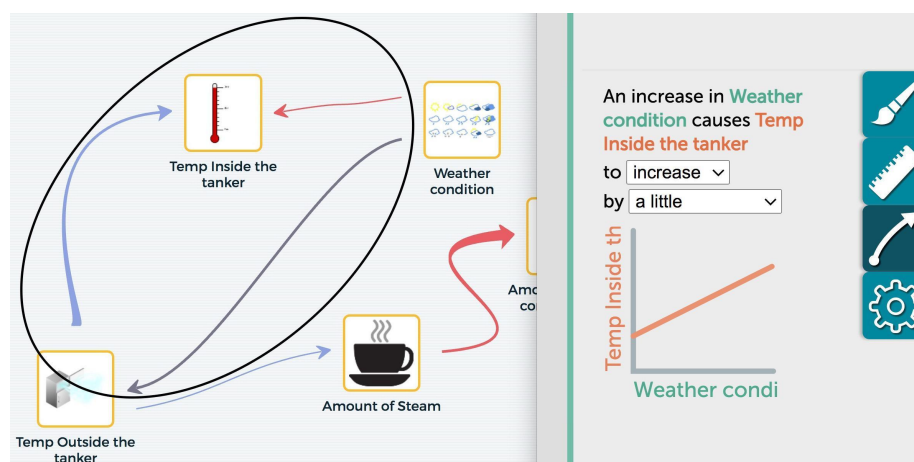
May: “Yeah, cold water and hot water don’t really mix.”

I: “So if I understand what you are saying there, then the blue arrow means that the temperature outside the tanker caused the temperature inside the tanker to cool down? Because the temperature outside got colder”

May: “Um, well, yes and no. I believe that with the collision of them both coming together and mixing, um, kind of like I think how tornadoes, kind of how cold and hot kind of make it [spins her fingers around each other], a tornado, go.... It was like, colliding against each other.”

Figure 5

Compensating Effects in Final Model of May and Her Partner



Weather condition *seems to have compensating effects on* Temp Inside the tanker.²

Although May’s thinking was still not completely clear, she used other terms of conflict as well, saying the molecules on the inside and outside of the tanker have to balance so that they are “not constantly like at each other’s throats or whatever, banging against each other.” During the unit, the students had investigated simulations of gas molecules colliding on both sides of a movable barrier, which could have been one source of her vivid imagery.

Student Issue and Questioning Strategy That Preceded the Response

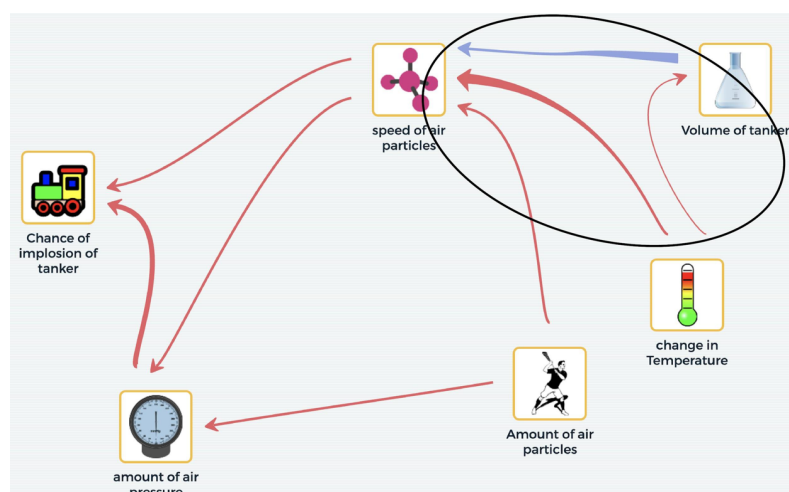
May exhibited multiple issues with reasoning, including non-canonical ideas. In her model, the effects of the two pathways between *Weather condition* and *Temp Inside the tanker* would largely cancel if the model were simulated, but the student appeared to see the effects as conflicting. This could have been a case of misapplying a conception she had learned elsewhere about tornadoes.

The interviewer questions were prompted by a causal pattern in the model that did not appear to make sense. In this instance, the interviewer began by asking about individual relationships. When the student gave an unclear explanation about the negative link between *Temp Outside the Tanker* and *Temp Inside the Tanker*, the interviewer probed the student thinking about that relationship. Although the interviewer's questions did not explicitly ask about a combination of relationships, the questions were inspired by, and focused on, an unexpected causal pattern, which then appeared to serve as a mutual visual referent for interviewer and student. We suggest this focus prompted May to shift from explaining individual relationships to explaining a pair of relationships. In doing so, she articulated her unexpected reasoning about the underlying molecular interactions.

Derek Revisited: Parallel Causal Chains, Compensating Effects (Patterns C and D)

Parallel causal chains can be used when one variable affects another in multiple ways through more than one causal mechanism. This pattern appeared in about a third of the final models analyzed, but it was not clear that the students were thinking of multiple mechanisms.

In this section, we return to Derek's introductory example. Here we look at two earlier models in his sequence because he provided a particularly clear example of unexpected reasoning that can underlie this structure. We first examine a single link paired with a causal chain in Derek's third model (Figure 6), then its transformation into parallel causal chains in his fourth model (Figure 7).

Figure 6*Third Model of Derek and His Partner*

The circled portion of the model, with compensating effects, is transformed into parallel chains in the next revision (in Figure 7).

Derek's third model (Figure 6) had compensating effects superficially reminiscent of May's. These appeared to predict that a *change in Temperature* would have two different effects on *speed of air particles*. Neither Derek nor his partner mentioned this in their written explanations. When examining their entire model revision sequence to prepare for Derek's final interview, we suspected that the issue was more similar to Mark's, a failure to consider the combined effects of the two-link causal chain leading through *Volume of tanker*. We assumed that by *change in Temperature* these students meant simply *Temperature*, and we doubted they intended the two-link chain to indicate that increasing temperature would cause particles to slow. We suspected their thinking about that chain was separate from their thinking about the single link between temperature and particle speed, and that they had not considered the compensating effects. Because Derek's first interview had occurred shortly before he and his partner constructed this model revision, we reviewed that transcript to gain a better idea of his understanding as it existed then, and we followed up during his final interview.

In that first interview, Derek had said he thought the volume affected the speed of the air particles because if they were in a smaller volume they would bounce around more.

(Ellipses denote repetitive utterances omitted for clarity.)

Derek: “We thought the *change in Temperature* directly affected the *speed of the air particles*. Like, if the temperature was hot, then we thought the air particles would go faster. . . . And then . . . we had *Volume* and we thought that it directly affected the *speed of the air particles* because if they’re in a smaller volume they’d bounce around more.”

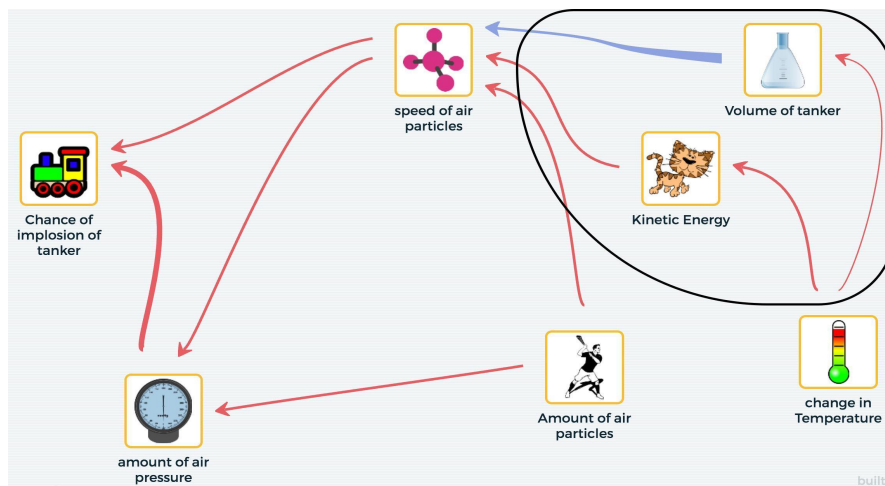
This reinforced our doubts that the students had intended the two-link chain to indicate that increasing temperature would cause particles to slow. It suggested that the relationships between *change in Temperature* and *speed of air particles* and between *Volume of tanker* and *speed of air particles* were being considered separately, at least at that time.

If Derek was thinking of two different effects of temperature on particle speed and doing so in terms of compensating effects (combining to produce little change) or effects that competed against each other (the way May thought of hot and cold water), he had not spoken in either of those terms in the first interview. He had elaborated on his thinking about the effect of *Volume of tanker* on *speed of air particles* by comparing it with another effect, that of *Amount of air particles* on *speed of air particles*.

Derek: “If there’s a little bit, then they’re bouncing into each other only a little bit, but if there’s a lot, then they’re bouncing into each other a lot and they keep, like, picking up the speed. . . . And we also thought the volume affected the speed of air particles, just like I explained with the amount of air particles.”

If air particles pick up more speed the more they bounce into each other, then putting the same number into a smaller container would increase the speed, as indicated in his model.

Later these students would remove some links from their model, citing a wish to avoid redundancy (Figure 4). But in the causal pattern involving temperature in Figure 6 we suspected something more than redundancy was involved. Their next model revision, along with comments in Derek’s final interview, reinforced this impression.

Figure 7*Parallel Causal Chains in Fourth Model of Derek and His Partner*

A causal mechanism has been added between change in Temperature and speed of air particles, resulting in parallel causal chains. Compare with Figure 6.

In their fourth model (Figure 7) Derek and his partner added a variable and a link to help explain how temperature affected molecular speed. They wrote, “We ... added a direct relationship between temperature inside the tanker and kinetic energy and also ... [between] kinetic energy and speed of air particles.” This created parallel causal chains.

Only one of these chains, the new two-link causal chain through *Kinetic Energy*, remained in their final model (Figure 4). In his final interview, Derek talked about this chain, “The higher the *Temperature*, the more *Kinetic Energy*, because particles would start to move faster, which means that *speed of air particles* would go up.” This is an accurate description of the chain and reflected what students had learned in class.

This is different from the ideas that the parallel causal chain in Figure 7 appeared to represent. Although this pattern was no longer present in their final model (Figure 4), in his final interview, both before and after his explanation of the *Kinetic Energy* chain, he provided explanations consistent with the *Volume* chain that was no longer there. Before the final model was brought up on screen, he said the following.

Derek: “If it’s cold, the air molecules will have a smaller volume inside the tanker and it’ll make a lower pressure—no, a higher pressure. And then if it’s . . . hotter, then the molecules will be more spread out and it will be a lower pressure.”

The first sentence could have been referring to condensation, but the second sentence was a surprise. Later, after Derek had talked through his model, the interviewer probed about the three-link chain between *change in Temperature* and *amount of air pressure* (Figures 4 and 7), asking, “What does your model predict about how changing the temperature of the gas, with everything else the same, how that will affect the pressure?” Derek reiterated that with increasing temperature the particles would get more spaced out and the pressure would go down, although he did not know if their model said that. A plausible interpretation is that the parallel pathways in Figure 7 represent two different ways this student thought about the phenomenon. Although he made some effort to give a consistent prediction for the effect of a *change in Temperature* on the system, his explanations in his final interview suggest that he still had not fully integrated these two ways of thinking.

This is an example of student thinking confounded in a way that would result in a model with compensating effects, and that is what we see. The model seemed to accurately reflect his thinking, the pattern in the model indicated a good place to probe, and the strategy of asking a question that forced him to consider both conflicting pathways together helped him clarify what he understood and was representing with the model.

Student Issues and Questioning Strategy That Preceded the Response

Derek’s third model (Figure 6) had two relationships chained in a way that did not make sense, reminiscent of Mark’s. However, the underlying issues may have been different. The pathway through *Volume of tanker* appeared to reflect Derek’s non-canonical understandings that collisions cause particle speed (and hence more collisions in a smaller space causes more speed) and that the amount of spacing between gas particles is the determining factor for the amount of pressure (regardless of particle speed).

Questions in both of Derek's interviews were inspired by patterns in the models that represented complex causal relationships we were not convinced the student intended. The question that elicited the clearest articulation of his (unexpected) reasoning was one that asked the student what his model would predict about a distal relationship, the effect on pressure of changing the temperature of the gas. His answer conflicted with the final model in front of him but was consistent with a now-deleted pathway from his previous model.

Discussion

We wondered whether classroom discourse and student artifacts were failing to reveal important reasoning behind student system modeling choices, and we wanted to explore how to elicit this reasoning. In this section, we draw from our analysis of interview data to describe a questioning strategy that appeared to support students' articulation of this thinking and we characterize the different student issues that surfaced when the strategy was used.

Questioning Strategy

In Stephens and Roderick (2019), we anticipated that each causal pattern in student models would suggest a different prompting question. However, from in-depth analysis of the questions asked during interviews we found that although productive questions were more varied than we anticipated, they all appeared to arise from a single strategy, that of encouraging a focus on distal effects. These questions had not arisen from knowledge of the patterns, which were identified later, but from our puzzlement over the meaning of those combinations of relationships in the models. These questions often, but not always, took the general form of "*What effect do you think Variable A has on Variable [C or D]?*" We hypothesize that drawing attention to combinations of effects helped students expand their focus from a single variable to a relationship, from a single relationship to a chain, or from a single chain to interacting chains. With each shift in focus, a further kernel of student reasoning was elicited and discussed.

The Role of Model Substructures

The causal patterns in Table 1 fall under Perkins and Grotzer's (2005) category of Multiple Linear Causality, which includes "[d]omino causalities where effects in turn become causes as in simple causal chains ... or branching patterns" (p. 126). They rate this category as the second simplest in six categories of increasingly complex causal interaction patterns. We suggest that teachers may need supports for noticing and facilitating discussions about them. In *SageModeler*, these patterns take the form of characteristic topological substructures, the simplest forms of which are in Table 1. In interviews, the substructures appeared to form mutual visual referents when questions focused on them. Once we had identified them, we found that we were able to recognize these substructures easily even within complicated models. These patterns are only somewhat more complex than individual relationships, and, we suggest, are accessible to many students if attention is directed to them. They appear to form a "mid-level pattern of complexity" comparable to that identified by Levy and Wilensky (2004) in the context of agent-based reasoning. Noticing substructures and asking questions about the distal relationships within them may be a practical strategy for teachers to use to elicit student thinking while the student is still actively involved in modeling. There is also the possibility that students themselves can direct their partner's attention to these substructures when they build a model together. Authors such as King (1998) have made powerful arguments about the effectiveness of scaffolded peer-to-peer interactions. Our findings are consistent with theirs; it appears that well-designed general questions can activate higher forms of cognition.

Student Reasoning

The reasoning that gave rise to combinations of relationships that did not make sense in the models could arise from several causes. We found that probing was required to uncover which ones were at work.

Challenges with modeling conventions were similar to those identified by Ogborn (1999) and Bliss (1994), e.g., trouble understanding the meaning of a negative (“change in opposite direction”) link or understanding that the effect of one variable is passed along to all variables interlinked to it. Many students did not realize that when there were multiple links pointing to a variable, the model would combine the effects.

Students often found it challenging to talk through cause and effect relationships in their models. In essence, Mark said that the bigger the tanker, the more air in the tanker, the higher the air pressure. The explicit prompts we initially tried did not work. Until his attention was drawn to the distal relationship in a different way, by pointing out a graph he had made of the relationship, he failed to notice the implication that an expanding volume would produce more pressure. Difficulties such as his may have stemmed from attempts to deal with multiple variables in ways described by Rozier and Viennot (1990) and Kautz et al. (2005), where students lost track of variables that were being held constant. Students also had trouble reasoning along causal chains that contained negative relationships, consistent with Ogborn (1999).

Some of the non-canonical ideas that emerged in the interviews were reminiscent of reasoning in thermodynamics reported by Rozier and Viennot (1990) and Kautz et al. (2005). However, May expressed an idea we had not encountered before, that “cold and hot water don’t really mix,” but would compete the way “cold and hot kind of make ... a tornado, go.” Derek thought that increasing the temperature of a gas would decrease the pressure because there would be more space between the molecules. None of these ideas had been in evidence in classroom discussions and were only hinted at in these students’ written explanations.

We found that a particular kind of explicit prompt was often needed to elicit student articulation of these issues. Questions focusing on distal relationships within substructures of mid-level complexity could reveal unexpected thinking about system modeling conventions,

causal reasoning, and/or non-canonical ideas. These correspond to three areas of knowledge Hmelo-Silver and Azevedo (2006) identified as necessary in order to engage in complex systems thinking and modeling.

Limitations and Implications

The examples were from the classes of a single teacher, although we observed these issues across teachers and units to varying degrees. Modeling and content difficulties could have been a result of failings of the design of the modeling tool and/or the unit, although the literature documents similar difficulties existing more broadly. Because the questioning strategy was identified after the interviews were conducted, further studies are warranted to determine the usefulness of the strategy when applied systematically in other contexts and by other people, including teachers and students. As the model substructures are content agnostic, they can be detected automatically, offering additional possibilities for teacher and student support and additional research avenues.

Conclusion

Unexpected combinations of relationships in student models led us to suspect there was underlying reasoning that had not been evident in classroom discussions or in students' written work. Once this reasoning was elicited, we found that many of the issues had been identified in the literature. However, the method we used to detect them is one we have not seen described. It focuses on causal patterns in the form of substructures of intermediate complexity that include distal effects that may not have been intended by the students. Questions about the interacting relationships producing those effects can reveal inchoate thinking and may be productive for teachers who want to introduce pre-college students to systems thinking and need specific supports for how to facilitate discussions about complex causality. Rather than viewing system models of novice modelers as accurate representations of their thinking, our results imply that it may be more useful to view them as a source of

questions to elicit underlying reasoning, which could be unexpected. We suggest that using a questioning strategy such as the one proposed here could help teachers encourage a shift in focus from individual system elements to more complex causal patterns, beginning with representations students themselves have constructed but may not fully understand.

¹ Variables separated by a causal chain of at least two links are *distally* related as opposed to *proximally* related.

² The gray arrow in the figure is a student-defined relationship. These were not a focus of instruction, although a few students wanted to try them. This one was not well-defined as positive or negative and had little effect on the behavior of the model, but did indicate a second causal pathway.

Acknowledgements. We thank the teacher who participated in this work, Erin Cothran.

Funding. This material is based upon work supported by the National Science Foundation under Grant Nos. DRL-1417809, DRL-1417900, DRL-1842035, and DRL-1842037.

Declarations

Disclaimer. The opinions expressed herein are those of the authors and not necessarily those of the NSF.

For more information, see <https://concord.org/our-work/research-projects/building-models/> and <https://concord.org/our-work/research-projects/multilevel-computational-modeling/>.

References

- Ben-Zvi Assaraf, O., & Orion, N. (2005). Development of system thinking skills in the context of Earth science system education. *Journal of Research in Science Teaching*, 42(5), 518–560.
- Ben-Zvi Assaraf, O., & Orion, N. (2010). System thinking skills at the elementary school level. *Journal of Research in Science Teaching*, 47(5), 540–563.
- Bliss, J. (1994). Reasoning with a semi-quantitative tool. In H. Mellar, J. Bliss, R. Boohan, J. Ogborn, & C. Tompsett (Eds.), *Learning with artificial worlds: Computer based modelling in the curriculum* (pp. 128–141). Falmer Press.
- Booth Sweeney, L., & Sterman, J. (2000). Bathtub dynamics: Initial results of a systems thinking inventory. *System Dynamics Review*, 16(4), 249–286.
- Brandstädter, K., Harms, U., & Großschedl, J. (2012). Assessing system thinking through different concept-mapping practices. *International Journal of Science Education* 34(14), 2147–2170
- Chi, M., de Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18, 439–477.
- Collins, A., Brown, J., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator*, 15(3), 6–11.

- Creswell, J. (2007). *Qualitative inquiry & research design: Choosing among five approaches*. Sage Publications.
- Fisher, D. (2018). Reflections on teaching system dynamics to secondary school students for over 20 years. *Systems*, 6(2), 12–30.
- Gilbert, J. K. (2004). Models and modelling: Routes to more authentic science education. *International Journal of Science and Mathematics Education*, 2(2), 115–130.
- Gilissen, M., Knippels, M.-C., & van Joolingen, W. (2020) Bringing systems thinking into the classroom. *International Journal of Science Education*, 42(8), 1253–1280.
- Hammer, D. (1996). More than misconceptions: Multiple perspectives on student knowledge and reasoning, and an appropriate role for education research. *American Journal of Physics* 64(10), 1316–1325.
- Hmelo-Silver C. E., & Azevedo, R. (2006). Understanding complex systems: Some core challenges. *Journal of the Learning Sciences*, 15(1), 53–61.
- Hmelo-Silver, C. E., Jordan, R., Eberbach, C., & Sinha, S. (2017). Systems learning with a conceptual representation: A quasi-experimental study. *Instructional Science*, 45(1), 53–72.
- Hogan K., & Thomas, D. (2001). Cognitive comparisons of students' systems modeling in ecology. *Journal of Science Education and Technology*, 10(4), 319–345.
- Hokayem, H., & Schwarz, C. (2014). Engaging fifth graders in scientific modeling to learn about evaporation and condensation. *International Journal of Science and Mathematics Education*, 12(1), 49–72.
- Hrin, T., Milenkovic, D., Segedinac, M., & Horvat, S. (2017). Systems thinking in chemistry classroom: The influence of systemic synthesis questions on its development and assessment. *Thinking Skills and Creativity*, 23, 175–187.
- Kautz, C., Heron, P., Loverude, M., & McDermott, L. (2005). Student understanding of the ideal gas law, Part I: A macroscopic perspective. *American Journal of Physics*, 73(11), 1055.
- King, A. (1998). Transactive peer tutoring: Distributing cognition and metacognition. *Educational Psychology Review* 10, 57–74.
- Komis, V., Ergazaki, M., & Zogza, V. (2007). Comparing computer-supported dynamic modeling and 'paper & pencil' concept mapping technique in students' collaborative activity. *Computers & Education*, 49(4), 991–1017.
- Kuhn, D. (2007). Reasoning about multiple variables: Control of variables is not the only challenge. *Science Education*, 91(5), 710–726.
- Kuhn, D., & Dean, D. (2004). Connecting scientific reasoning and causal inference. *Journal of Cognition and Development*, 5, 261–288.
- Lee, M. Y. (2021). The potential relationship between clinical interview skills and mathematics teacher noticing: An exploratory study. *International Journal of Science and Mathematics Education*, 19(4), 793–814.
- Lemke, J. L. (1990). *Talking science: Language, learning, and values*. Ablex.
- Levy, S. T., & Wilensky, U. (2004, April 12-16). *Making sense of complexity: Patterns in forming causal connections between individual agent behaviors and aggregate group behaviors* [Paper presentation]. American Educational Research Association (AERA) Annual Meeting, San Diego, CA, United States.
- Metcalfe, S. J., Krajcik, J., & Soloway, E. (2000). Model-It: A design retrospective. In M. J. Jacobson & R. B. Kozma (Eds.), *Innovations in science and mathematics education:*

- Advanced designs for technologies of learning* (pp. 77–115). Lawrence Erlbaum Associates.
- Miller, R., Ogborn, J., Briggs, J., Brough, D., Bliss, J., Boohan, R., Brosnan, T., Mellar, H. & Sakonidis, B. (1993). Educational tools for computational modelling. *Computers & Education* 21(3), 205–261.
- Nassaji H., & Wells, G. (2000). What's the use of 'triadic dialogue'? An investigation of teacher-student interaction. *Applied Linguistics* 21(3), 376–406.
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.
- NGSS Lead States. (2013). *Next generation science standards: For states, by states*. National Academies Press.
- Nguyen, H., & Santagata, R. (2021). Impact of computer modeling on learning and teaching systems thinking. *Journal of Research in Science Teaching*, 58(5), 661–688.
- Ogborn, J. (1999) Modeling clay for thinking and learning. In W. Feurzeig & N. Roberts, (Eds.), *Modeling and simulation in science and mathematics education* (pp. 5–37). Springer.
- Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching* 41(11), 994–1020.
- Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books.
- Perkins, D., & Grotzer, T. (2005). Dimensions of causal understanding: The role of complex causal models in students' understanding of science. *Studies in Science Education*, 41(1), 117–166.
- Pimentel D., & McNeill, K. (2013). Conducting talk in secondary science classrooms: Investigating instructional moves and teachers' beliefs. *Science Education* 97(3), 367–394.
- Plate, R. (2010). Assessing individuals' understanding of nonlinear causal structures in complex systems. *System Dynamics Review*, 26(1), 19–33.
- Rozier, S., & Viennot, L. (1990). Students' reasoning in thermodynamics. *Tijdschrift voor Didactiek der B-wetenschappen*, 8(1), 3–18.
- Russ, R., Coffey, J., Hammer, D., & Hutchison, P. (2009). Making classroom assessment more accountable to scientific reasoning: A case for attending to mechanistic thinking. *Science Education* 93(5), 875–891.
- Sommer, C., & Lucken, M. (2010). System competence—Are elementary students able to deal with a biological system? *NorDiNa* 6(2), 125–143.
- Stephens, L., & Ki, L. (2017, April 22-25). *Explanations and relationships in students' mental and external models* [Paper presentation]. Annual Meeting of the National Association for Research in Science Teaching (NARST), San Antonio, TX, USA.
- Stephens, L., & Roderick, S. (2019, August 26-30). Making sense of students' static equilibrium models. In T. Bielik (Chairperson), *International perspectives on modeling and systems thinking* [Symposium]. European Science Education Research Association (ESERA) 13th Conference, Bologna, Italy.
- Stratford, S., Krajcik, J., & Soloway, E. (1998). Secondary students' dynamic modeling processes: Analyzing, reasoning about, synthesizing, and testing models of stream ecosystems. *Journal of Science Education and Technology*, 7(3), 215–234.

- Stroup, W., & Wilensky, U. (2014). On the embedded complementarity of agent-based and aggregate reasoning in students' developing understanding of dynamic systems. *Technology, Knowledge and Learning*, 19, 19–52.
- Tripto, J., Ben-Zvi Assaraf, O., Snapir, Z., & Amit, M. (2016). The 'What is a system' reflection interview as a knowledge integration activity for high school students' understanding of complex systems in human biology. *International Journal of Science Education*, 38(4), 564–595.
- van Zee E., & Minstrell, J. (1997). Reflective discourse: Developing shared understanding in a physics classroom. *International Journal of Science Education* 19(2), 209–228.
- Weiland, I., Hudson, R., & Amador, J. (2014). Preservice formative assessment interviews: The development of competent questioning. *International Journal of Science and Mathematics Education* 12(2), 329–352.
- Wilensky, U. (2003). Statistical mechanics for secondary school: The GasLab Modeling Toolkit. *International Journal of Computers for Mathematical Learning*, 8(1), 1–41.
- Yin, Y., Vanides, J., Ruiz-Primo, M.A., Ayala, C.C., & Shavelson, R. J. (2005). Comparison of two concept-mapping techniques: Implications for scoring, interpretation, and use. *Journal of Research in Science Teaching*, 42(2), 166–184.
- Yoon, S. A., Goh, S. E., & Park, M. (2018). Teaching and learning about complex systems in K–12 science education: A review of empirical studies 1995–2015. *Review of Educational Research*, 88(2), 285–325.