The logic of Babel: Causal reasoning from conflicting sources

Matthew W. Easterday[1], Vincent Aleven, Richard Scheines
Human-Computer Interaction Department, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh PA, 15213

Abstract Ill defined problems lack structure partially because there is no agreed upon way of representing the problem. In this follow-up study, we examine how diagrams help students learn to analyze policy arguments. Our previous work asked students to predict the effect of a policy intervention based on testimonies from conflicting sources, and showed that teaching students a formal, diagrammatic procedure improved students' predictions. In this study we looked at *how* students and experts use diagrams so that we could a) identify errors in student reasoning and b) start to develop a cognitive model of construction and interpretation of causal diagrams. We thus conducted an informal protocol analysis on how 4 novices and 3 experts solved causal reasoning problems using 1) text, 2) text and a correct diagrammatic representation, and 3) text with a diagramming tool. We found that many of the errors in causal reasoning stemmed not from the difficulty of using diagrams per se, but from conflicts of background knowledge with the provided testimonies. Some participants demonstrated a diagram confirmation bias, i.e. they reinterpreted the diagram syntax to reach a conclusion more consistent with their beliefs. Other participants made arguably normative "errors," i.e. they correctly interpreted the sources' claims in the testimony, but judged their own knowledge to be more credible. Allowing students to apply arbitrary background knowledge poses a problem for intelligent tutors that require a fully specified problem space. We conclude that tutors may be able to distinguish between confirmation bias and normative uses of background knowledge by asking students to explicitly add their background knowledge to the diagram.

Introduction

A rational, participatory democracy depends on an informed citizenry, one that can reason about the conflicting policy claims of multiple sources (Gore 2007). For example, if the U.S. Secretary of Defense states that "preemption will decrease a weapons of mass destruction (WMD) threat," while a policy analyst questions the claim, asserting instead that "international sanctions and foreign aid are a better way to reduce anti-American sentiment," and weapons experts argue that "rogue regimes with nuclear material are likely to increase proliferation," citizens must be able to weigh the various claims and judge the likelihood that a suggested policy (preemption) will lead to the desired outcome (a decreased WMD threat)–they must make logical judgments from "babel."

Voss (2005) describes seven features of policy problems that make them ill structured including their: lack of a clear goal state, having no objectively correct answer, etc. We are especially interested in Voss's third feature: the lack of an agreed upon strategies for representing policy problems. Unlike algebraic word problems that have

1 The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B040063 to Carnegie Mellon University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

well-defined procedures for converting words into equations (the formal representation), and for solving the equations, there is no agreed upon representation of a policy problem. In this paper we consider the question: how *should* solvers represent ill defined policy problems in order to make inferences about the probable effect of a policy intervention. For example, if a citizen decides that the policy analyst and the weapons experts are credible, how should he then judge the likelihood that preemption will lead to a decreased WMD threat? If, as Simon (1981) conjectured, "…solving a problem simply means representing it so as to make the solution transparent," then the representation issue is not only central for policy, but for ill defined problems as well.

Work on external representations suggests that diagrams might improve reasoning (Larkin, & Simon 1987; Ainsworth, 2006; Harrell, 2004; Scaife & Rogers, 1996, Cox, 1999, Novick & Hurley, 2001, Mayer & Moreno, 2002, Bauer & Johnson-Laird 1993, Pinkwart, Aleven, Ashley & Lynch, 2006), but not which specific diagrams should be used for policy problems, or how to design them. Since arguments about policy often hinge on causal assumptions or inferences, we are focused on the benefits to learners of causal diagrams, a type of diagram that maps out causal relations claimed to exist within a topic area (e.g., the effects of preemption or international sanctions). Such diagrams and the associated causal theory (Spirtes, Glymour, Scheines 2000) lend a significant amount of structure to the domain, but do not render it well-defined. Even after a solver commits to a causal representation, there is no single correct way to represent the causal factors or to choose the grain size at which to represent those factors.

In prior work, we showed that causal diagrams can be helpful to students as they learn to interpret brief (experimenter-written) policy texts (Easterday, Aleven & Scheines, 2007). We found that providing students with a causal diagram that summarizes a particular policy text helps them do better in interpreting that text: students take advantage of the diagram to make better predictions about the effects of policy interventions described in text. We also found that having students practice constructing diagrams for policy texts supports learning how to interpret new texts, even when (as would be the case for a newspaper article) the new texts are not accompanied by a causal diagram. Thus, the previous study implies that an intelligent tutoring system that helps students in constructing causal diagrams will contribute to their skill in reasoning about policy. In order to develop intelligent tutors it is important to understand students' strategies for analyzing a policy texts and constructing diagrammatic representations of the texts as well as the difficulties that students experience in that process. Addressing these questions will yield a greater understanding of argumentation in ill-defined domains and the use of diagrammatic representations in that context.

Previous research has demonstrated a variety of errors (Kuhn 1991) and expert-novice differences (Voss 1983) in policy reasoning. These studies however were not focused on the use of diagrams. Following in their footsteps, the goals of the follow-up study were to gain insight into how students and experts used diagrams to make inferences about the effects of a policy in order to a) identify errors in reasoning, and b) to inform the design of a cognitive tutor.

1. Method

Task and Intervention

As in the previous study, we gave participants short, fictional, policy texts (Figure 1).

Childhood obesity is now a major national health epidemic. A number of facts are widely agreed upon by the public and scientific community: exercise decreases obesity, and eating junk food increases obesity. It's also clear that people who watch more TV are exposed to more junk food commercials.

Parents for Healthy Schools (PHS), an advocacy group which fought successfully to remove vending machines from Northern Californian schools, claims that junk-food commercials on children's television programming have a definite effect on the amount of junk food children eat. In a recent press conference, Susan Watters, the president of PHS stated that "...if the food companies aren't willing to act responsibly, then the parents need to fight to get junk food advertising off the air."

A prominent Washington lobbyist Samuel Berman, who runs the Center for Consumer Choice (CCC), a nonprofit advocacy group financed by the food and restaurant industries, argues that junk food commercials only "influence the brand of food consumers choose and do not not affect the amount of food consumed." While Mr. Berman acknowledges that watching more TV may cause people to see more junk food commercials, he remains strongly opposed to any governmental regulation of food product advertising.

Recent studies by scientists at the National Health Institute have shown that watching more TV does cause people to exercise less.

Figure 1. Policy text on obesity.

We then asked participants to answer questions like: "According to the combination of claims made by the CCC and NHI, will making kids watch less *TV*, decrease childhood *obesity*?" According to the procedures taught in the experiment, students should notice that the CCC denies the effect of TV on obesity *through commercials*, and the NHI claims an effect of TV on obesity *through lack of exercise*, so the correct answer is *yes*, TV will affect obesity (according to the claims of the CCC and NHI). These questions simulate the difficult task of assembling the claims of multiple sources to predict the likely affect of a policy intervention. Note that real policy problems like terrorism or the environment are far more complex, however students find even these simple texts to be quite challenging.

Participants received policy information in one of the following forms:

1. **Text (only)** in which the case studies was presented in text only (Figure 1).
2. **Diagram (+ text)** in which the case study was presented as text accompanied by a correct, diagrammatic representation of the case study (Figure 2).
3. **Tool (+ text)** in which the case study was accompanied by a computer tool with which participants could construct their own diagrams.

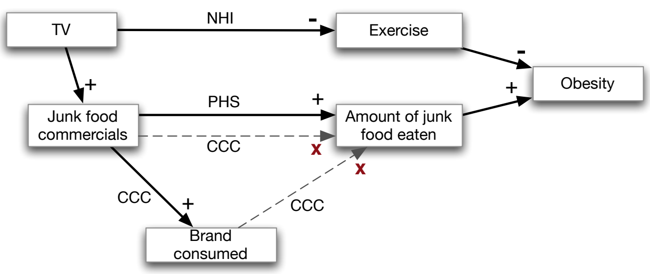

Figure 2. A causal diagram representing the case study on obesity. Boxes represent causal variables, and arrows represent either positive (+), negative (-), or no (x) influence of one variable on another. An annotation on the arrow (e.g. PHS) identifies the source making the causal claim.

Note that to solve the question about TV increasing obesity (according to the CCC and NHI) using the diagram in Figure 2, participants should notice the arrow from TV to exercise labeled "NHI", and the unlabeled arrow from exercise to obesity (representing common knowledge), as they are taught during the experiment.

Participants and setting

In this protocol analysis study, we examined 3 "experts" from the Carnegie Mellon University (CMU) Philosophy Department, all of whom have a PhD in philosophy and who have conducted original research on causal reasoning, and 4 CMU student "novices." All participants were offered $20, however all experts declined payment.

Research design
In this protocol analysis, participants were asked to "talk aloud" as they completed an on-line lesson on causal reasoning. We assigned one novice and one expert to each of the text, diagram and tool conditions, and a second novice to the diagram condition to collect more data. After a pretest involving a case study on the environment similar to Figure 1, each participant received a 4 page, interactive, on-line tutorial on causal reasoning that also included diagrams like Figure 2 for the diagram and tool groups. To make the training for the text as close to identical as possible, every diagrammatic explanation in the diagram/tool training was matched by an equivalent prose explanation in the text training. Following training, we tested all participants on the case study in Figure 1, presented as text only to the text participants, as text with a correct diagram to the diagram students, and as text with a tool to the tool participants.

Data collection and analysis
To measure performance, participants were tested on 10 multiple choice, causal questions (e.g. "According to the PHS, will making kids exercise more reduce the number of junk food commercials they watch?"). Participants could answer either: a) *yes* there would be a causal effect (e.g. making kids exercise more would reduce the number of junk food commercials they watch), b) *no* there would be no causal effect, or c) *inconclusive* the sources explicitly disagree about the causal effect.

To capture process information, participants' speech and on-screen behavior were recorded with a screen capture program. Because we intend to build tutors for automated knowledge tracing and instruction, we did not create a coding manual or use multiple coders to analyze the video; we instead used the screen recordings to identify processes and errors to inform the design of a cognitive model.

2. Results

Text (Novice 1, Expert 1)
Both the novice and expert in the text condition performed quite poorly. Novice 1 scored 20% while Expert 1 scored 0% on the first half of the questions, after which Expert 1 ended the experiment stating: "My brain is fried." While Expert 1's performance seems abysmal, recall that in this condition, Expert 1 did use his standard tool (a causal diagram) and, unlike Novice 1, Expert 1 realized the difficulty of completing the task without a diagram. These scores compare with a chance score of 33% (guessing randomly between the 3 options of yes/no/inconclusive), and the average score of 41% on the same condition in the previous study. This performance underscores the difficulty of reasoning about even simple causal systems using text alone. We take the ubiquity of causal diagrams in causal reasoning research, the poor results of the text group in the previous study, and the difficulty of using text by both the novice and expert in this study, as an indication to focus our future efforts on diagram use.

Tool + Text (Novice 2, Expert 2)
In the previous study, students who were given case studies as text accompanied by a diagramming tool scored an average of 40%, performing no better than the text group. Given the poor performance of this diagram construction group in the previous study, we expected Novice 2 to have difficulty with diagram construction. In fact, both Novice 2 and Expert 2 made better diagrams than most observed in the previous study.

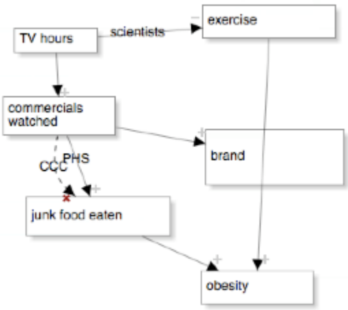

Figure 3. Expert 2's diagram. The experts diagram is virtually correct.

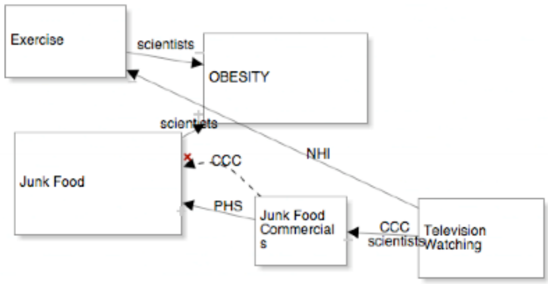

Figure 4. The Novice 2's diagram. Note omitted "brand" variable and the mislabeled links from junk food to exercise.

Despite making relatively good diagrams, small errors in diagram construction sometimes lead to relatively large errors in interpretation. For example, by mislabeling the two arrows pointing to obesity, Novice 2 might answer every question about obesity, (the focus of the case study,) incorrectly. While both their diagrams contained errors, Expert 2's diagram (assuming it was used correctly to answer the test questions) would have lead to the correct answer on 100% of the questions, whereas Novice 2's diagram would have lead to the correct answer on 20% of the questions.

However, both participants often used their diagrams incorrectly. If we grade them on their ability to use the diagram they constructed, (in an algebra word problem where students must translate the problem into an equation and solve the equation, this would be like grading students based only on whether or not they solved the equation correctly, even if the equation they started with was incorrect), Expert 2 used the diagram correctly on only 60% of questions, Novice 2 on 20%. If we grade them based on whether they got the right answer, we find that (coincidentally,) Expert 2 answered only 60% of the questions correctly, Novice 2, 20%.

Diagram + Text (Novice 3 & 4, Expert 3)

During training, participants in the diagram and tool conditions were taught a procedure for interpreting diagrams. We thought that giving students a correct diagram in addition to the text would improve their reasoning. Although the previous study showed that providing a correct diagram did indeed improve performance, this study shows that participants' background knowledge sometimes overrules the conclusions implied by the diagram, and sometimes leads to a reinterpretation of the diagram syntax to reach a conclusion more consistent with background beliefs.

Participants overrule the diagram with background knowledge when they make:

- *Override errors*, where the reasoner correctly reads the graph, but decides that their background knowledge is more credible. This can be normative if the reasoner: a) makes separate and correct predictions about the both the evidence provided and their beliefs, and b) explicitly claim their beliefs to be more credible than the evidence provided.
- *Speculation errors*, such as adding information to the diagram about what a source *would say*, given what that source has already said.

Participants selectively reinterpret the diagram (confirmation bias) when they make:

- *Diagram interpretation errors*, such as confusing observation and intervention, i.e. believing that an arrow showing that A causes B, can also mean that B causes A.
- *False uncertainty errors*, such as interpreting a lack of an arrow by a source as indicating that "we don't know what the source thinks" instead of that "the source makes no claim."

And sometimes participants simply interpret the diagram incorrectly when they make:

- *Combination errors,* where the relevant paths are noticed, but not combined correctly to make the proper inference.
- *Impasse errors,* such as giving up on the diagram (and text) altogether.

Question	Type of error		
chain (correct answer: "yes")	Novice 3	Novice 4	Expert 3
a According to the NHI, will making kids exercise more reduce childhood obesity?	+	+	+
b According to the NHI & CCC, will making kids watch less TV decrease childhood obesity?	combination?	+	combination?
none (correct answer: "no")			
e According to the PHS will watching TV cause children to exercise less?	+	+	uncertainty & speculation
f According to common knowledge, will making children watch less TV decrease childhood obesity?	+	uncertainty	+
common cause (correct answer: "no")			
g According to the NHI, will making kids exercise more reduce the number of junk food commercials they watch?	interpretation	interpretation	+
h According to the NHI, will reducing the number of junk food commercials children watch reduce childhood obesity?	+	uncertainty	uncertainty
common effect (correct answer: "no")			
i According to common knowledge, will making kids exercise more reduce the amount of junk food they eat?	+	impasse	+
j According to the PHS, will making kids exercise more reduce the number of junk food commercials they watch?	+	impasse	override

Table 1. Errors made by diagram participants. The first column shows each of questions asked on the test, grouped by the underlying causal structure of the answer.. Cells indicate that the question was answered correctly ("+"), or the type of error made. No errors were made on questions *c* & *d* (not shown).

Diagram + Text: Novice 3

Novice 3 had the best performance, answering 80% of questions correctly, (as compared with an average score of 49% for diagram students in the previous experiment, putting Novice 3 in the top 15 percentile of that group). For the most part, Novice 3 did not seem to reference background knowledge at all, but seemed to consistently apply the diagram interpretation procedure to each question. To the extent that Novice 3 ignored his background knowledge, he fit the pattern of a ***diligent novice,*** executing the procedure as instructed.

Diagram + Text: Novice 4

Novice 4 answered 50% of the questions correctly, much closer to the average score of 49% observed in the previous study. We could characterize Novice 4's behavior as including far greater interference from background knowledge:

1. On questions *f* and *h*, Novice 4 made *uncertainty errors*, concluding that if there are no arrows (path) between the relevant variables, then it is inconclusive whether the source would say one variable would affect the other. Note that in the diagram interpretation procedure, participants were instructed that if there are no arrows, the source would *not* claim that one variable would affect the other, i.e. the answer is *no*. On question *f*, Novice 4 explains:

   "it doesn't say anything on here... I can't tell from there, so from looking at that, that would be inconclusive..."

   ...and on question *h* Novice 4 explains:

   "it doesn't say anything about junk food commercials, so that would be inconclusive."

Novice 4 did not consistently apply this "no path means inconclusive" reasoning however. In fact, Novice 4 inferred that the answer was *inconclusive* when her background knowledge (that TV does affect obesity) contradicted the correct answer of *no*:

> "I would assume that if you're watching TV you're not playing…that would lead to less children being obese."

The quotes suggest that Novice 4 wanted to answer *yes* according to her background knowledge, and selectively reinterpreted the meaning of an absence of an arrow when the correct interpretation contradicted her belief. After answering question *f,* we asked why she chose *inconclusive* rather than *no*, to which she responded:

> "…my feeling is to go for *yes*, so I kind of compromised and went for inconclusive."

2. Novice 4 (and Novice 3) also confused observation with intervention on question *g*, incorrectly interpreting an arrow from TV to exercise as also meaning that exercise decreases TV, a *diagram interpretation error*:

   > "Well without looking at that I would say 'yes', but looking at this…so kids are exercising more, then they watch less TV, which means they have, watch less junk food commercials. But the question is…'will making children exercise more, reduce the number of commercials they watch'. I don't know about reading the graph backwards, its confusing. Well I'm going to say 'yes'."

   Note that the novice did not regularly infer that an arrow denoting that A causes B also means that B causes A. Novice 4 only made that error when it was consistent with her background knowledge, as can be seen in the above quote.

Unlike Novice 3, Novice 4 used a far greater amount of background knowledge, which unfortunately seemed to hurt performance. Although Novice 4 performed better than participants in the text condition, Novice 4 was not able to reconcile the diagram with her background knowledge, thus undermining the usefulness of the diagram. The reinterpretation of the diagram syntax to reach conclusions consistent with one's belief might be thought of as a kind of *diagrammatic confirmation bias.* Novice 4's behavior seems closer to that of a **confused novice** which is more representative of students in the previous study.

Diagram + Text: Expert 3

Expert 3 made more errors than expected, with an overall performance of 60%. While Expert 3, like Novice 4 made heavy use of background knowledge, Expert 3 seemed better able to separate his background knowledge from the predictions of the diagram, treating the two as separate entities not necessarily requiring reconciliation.

1. Expert 3's first error, on question *j*, was an *override error.* Expert 3 explicitly recognized a difference between his background knowledge and the claims of the diagram, and then explicitly chose to go with his background knowledge:

   > "Naturally I would assume that the PHS people would say "yeah it will reduce the number of junk food commercials they watch" because in fact, this guy up here, I think most people would think is actually a, uh, uh, goes both ways.... However, I'm supposed to answer the question based on what's been given to me so far... So I'm going to say the answer I'm supposed to give is 'no', but quite frankly, well you know what, I'm going to give the answer I think is right given the sorts of things I've got here, which is that its actually inconclusive. "

2. Expert 3 also made a *speculation error* on question *e,* where, given the fact that the PHS wants to decrease junk food advertising, Expert 3 inferred that the PHS would also agree with the NHI that TV would decrease exercise (note this would

be a reason to make kids watch less TV and therefore less junk food commercials). This speculation error was combined with an uncertainty error:

> "Well I'm willing to bet the PHS would absorb… well it's inconclusive, we don't know what the PHS thinks, we aren't given any context. ...So I'm going to say inconclusive, because I was not given that piece of information. Moreover, I think the PHS would presumably accept those kind of studies."

Like Novice 4, Expert 3 often applied his background knowledge to the problem. Unlike Novice 4, Expert 3 did not seem to be as confused by the diagram, as shown by his lack of diagram interpretation and impasse errors. When the diagram did not match his beliefs, he explicitly stated that the diagram is wrong and that he chose to rely on his background knowledge. For this reason, we characterize Expert 3 as a **truculent expert**. It may be that Expert 3 has more robust diagram reading skills than Novice 4, which would prevent him from questioning the implications of the diagram, and move on to the question of whether or not the diagram is correct.

Diagram + Text: summary
To clarify the differences between the diligent novice, truculent expert, and confused novice, we characterize the three patterns in Figure 6.

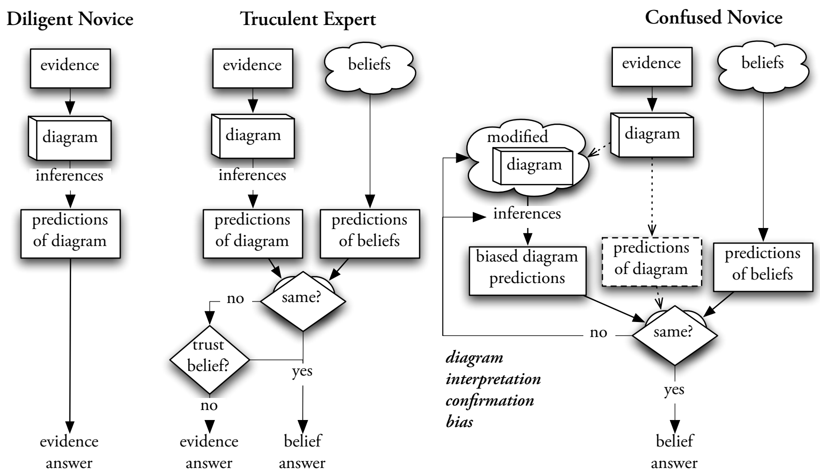

Figure 6. Three different patterns of reasoning.

3. Implications: Tracing background knowledge with tutors

The protocol data makes it clear that the task and measures developed in the previous study cannot detect whether a participant is making an incorrect statement as a truculent expert behaving normatively, or as a confused novice exhibiting confirmation bias. Redesigning the task and measures so that students must make their background knowledge explicit will allow us to monitor how they are using background knowledge. Once we can detect how they are using background knowledge, we can both evaluate their performance, and provide better tutoring. To detect whether students are applying their background knowledge normatively, we could do the following:

1. Ask students to make three inferences, one based on the diagram, one based on their background knowledge, and one based on both.

2. Use an editable (rather than fixed) diagram that allows students to add arrows (but not variables) representing their background knowledge to the diagram.
3. If measures 1 and 2 show that students have not made the correct inference based on the diagram or their background knowledge, we can ask them to highlight the arrows on which they based their decision. Then, for each highlighted, (or relevant, unhighlighted) arrow, we can ask whether the student: a) wants to add/reject a causal relation based on their own knowledge, b) speculate that a source would add/reject a causal arrow, or c) did not notice the arrow.

This procedure would allow us to retrospectively detect each of the errors observed in the protocol analysis. Figure 7 shows where the various interpretation errors arise during diagram interpretation, and how the proposed measures should be able to distinguish between normative uses of background knowledge, and confirmation bias.

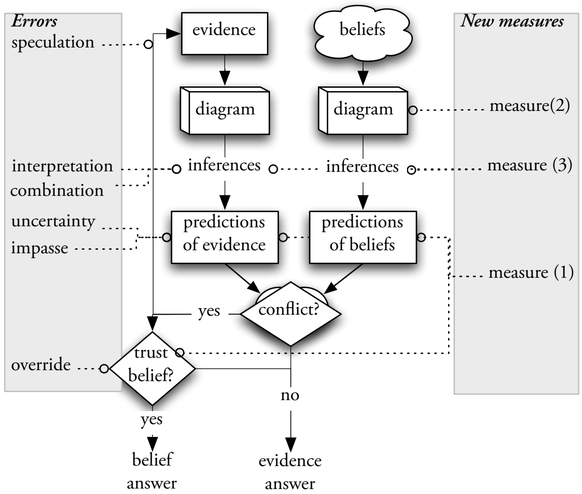

Figure 7. Errors of diagram interpretation and measures to detect them.

4. Discussion

This follow up study to Easterday, Aleven and Scheines (2007) showed that students' errors with diagrammatic representations stem not so much from the difficulty of the diagram construction or interpretation procedures per se, but rather the way in which the procedures conflict with students' background knowledge and informal reasoning.

Educational research tells us that in all domains, teachers (and tutors) "...must draw out and work with the preexisting understandings that their students bring with them." (NAS, 2000, p. 19). While background knowledge affects how students solve problems in general, it plays a more complicated role in ill defined problems. Simon (1973) describes a system for solving ill structured problems as:

> ...a combination of a GPS, which at any given moment finds itself working on some well structured subproblem, with a retrieval system, which continually modifies the problem space by evoking from long-term memory new constraints, new subgoals, and new generators for design alternatives. (p. 192).

...as opposed to "...bringing all of the potentially relevant information in long-term memory together once and for all at the outset, to provide a well structured problem space that does not change..." (p. 192). The fact that people use background knowledge in ill defined problems to continuously modify the problem state undermines cognitive tutors that rely on a fully specified problem state. How can a tutor trace knowledge and provide feedback in problems where the student is allowed, even encouraged, to apply background information not known to the tutor?

The study showed that background knowledge not only plays a role, but often manifests itself as a kind of *diagrammatic confirmation bias*. Kuhn (2005) describes instances in which students fail to *coordinate theory and evidence*, i.e. misinterpret a given set of facts (evidence) in order to support preconceived beliefs (theory). Kuhn rightly suggests not that students should ignore background knowledge, but rather that they should make judgments about evidence and background knowledge separately, so that they may compare the two (p. 72). Kuhn's admonition suggests that tutors for ill defined problems should not ask students to "[leave] their common sense at the door," but rather trace how students correctly (and incorrectly) apply their background knowledge. Tutors for ill defined problems do not have the luxury of assuming that, as in an algebra problem, the initial given facts determine a unique solution because in ill defined problems, students' background knowledge can sometimes trump the given facts.

Looking toward the future challenges in this line of research, this study suggests that tutors for policy argument must be able to monitor when and how students apply their background knowledge.

References

Ainsworth, S. E. (2006). DeFT: A conceptual framework for learning with multiple representations. *Learning and Instruction*, 16(3), 183-198.

Bauer, M. I., & Johnson-Laird, P. N. (1993). How diagrams can improve reasoning. *Psychological Science*, 4(6), 372-378.

Cox, R. (1999) Representation construction, externalised cognition and individual differences. *Learning and Instruction*, 9(4), 343-363.

Easterday, M. W., Aleven, V., & Scheines, R. (2007). 'Tis better to construct or to receive? The effects of diagram tools on causal reasoning. *Proceedings of the 13th International Conference on Artificial Intelligence in Education.*

Gore, A. (2007). *The assault on reason*. New York: Penguin Press.

Harrell, M. (2004). *The improvement of critical thinking skills in What Philosophy Is* (Tech. Rep. No. CMU-PHIL-158). Pittsburgh, PA: Carnegie Mellon University, Department of Philosophy.

Kuhn, D. (1991). *The skills of argument*. Cambridge, MA: Cambridge University Press.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11, 65-99.

Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction*, 12(1), 107-119.

National Research Council. (1999). *How people learn: Bridging research and practice*. Washington, DC: National Academy Press.

Novick, L. R., & Hurley, S. M. (2001). To matrix, network, or hierarchy: That is the question. *Cognitive Psychology*, 42(2), 158-216.

Pinkwart, N., Aleven, V., Ashley, K., & Lynch, C. (2006). Toward legal argument instruction with graph grammars and collaborative filtering techniques. *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*, 227-236.

Scaife, M. & Rogers, Y. (1996) External cognition: how do graphical representations work? *International Journal of Human-Computer Studies*, 45(2), 185-213.

Simon, H. A. (1973). The structure of ill structured problems. *Artificial Intelligence*, 4(3), 181-201.

Simon, H. A. (1981). The sciences of the artificial (2nd ed.). Cambridge, MA: MIT Press.

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). Cambridge, MA: MIT Press.

Voss, J. F. (2005). Toulmin's Model and the solving of ill-structured problems. *Argumentation*, 19(3), 321-9.

Voss, J. F., Tyler, S. W., & Yengo, L. A. (1983). Individual differences in the solving of social science problems. In R. F. Dillion, & R. R. Schmeck (Eds.), Individual differences in cognition (pp. 205-32). New York: Academic Press.