

Some Thoughts on Using Computers to Teach Argumentation

Kevin D. Ashley

Professor of Law and Intelligent Systems
Senior Scientist, Learning Research and Development Center
University of Pittsburgh

Thank you to:

Vincent Aleven

Assistant Professor, Human-Computer Institute
Carnegie Mellon University

Collin Lynch

ISP, University of Pittsburgh

Niels Pinkwart

Assistant Professor
Technische Universiteit Clausthal

Research sponsored by NSF Award IIS-0412830

Argument

- Given AI's focus on computational models of argumentation, it's important to develop techniques for evaluating the models.
- One way is to assess if the models help students learn argumentation skills.
- Research on computer-supported instruction with argument diagrams has shown primarily only suggestive evidence of success.
- Toulmin structures and other argument models and visual representations that focus on linking data and claims through warrants may not be ideal.
- Our visual representation in the LARGO ITS focuses instead on **argument strategies and processes for reasoning about the warrants**.
 - Even so, our evidence of learning is also mixed.
- I suggest some explanations for the mixed results of computer-supported instruction in argumentation and offer some alternative approaches.
- Finally, I report evidence of an emerging use of argument diagrams as **diagnostic** evidence of students' understanding of argumentation.

Outline

- Why teach argumentation, especially for ill-structured problems?
- Example ill-structured problem:
 - Arguments, argument models and argument diagrams.
- What do students need to learn about argumentation?
- Why teach argumentation with computer-supported argument diagrams?
- Does instruction with computer-supported argument diagrams really work?
- What some argument diagrams are missing:
 - [dis]advantages of Toulmin diagrams
- Our approach:
 - LARGO teaches arguing with hypotheticals
 - Report of a series of evaluations
- Why don't argumentation ITSs work better? What can be done?
- One role for argument diagrams in instruction: diagnostic tools
 - How diagnostic are LARGO diagrams?
 - Why diagnostic argument diagrams matter
- Future plans and conclusions

Why teach argumentation?

To make better reasoners!

“[A] good reasoner should be able to

- **generate arguments**, providing supportive reasons to the claims that are made....
- **consider arguments counter** to his or her argument and be able to refute them or to re-evaluate one’s own position in reference to them....
- use qualifiers that restrict the range of an argument....
- provide backing to an argument
- develop an argument structure consisting of ... inter-related arguments....
- **persuade via the strength of the arguments** provided....
- use various types of arguments (e.g., argument from consequence, argument by definition....)
- **evaluate arguments**, using the [relevant] criteria..” (Voss & Means, 1991, p. 342)

Argumentation and Ill-Structured Problems

“Ill-structured problems usually have:

- 1) *Vaguely stated goal* that requires analysis ... to define the particular issue clearly.
- 2) *Constraints ... not in the problem statement*; instead, the solver needs to retrieve and examine the constraints ... during the solving process.
- 3) [Solvers frame] ill-structured problems ... in different ways, according to [their] knowledge, beliefs, attitudes, and other factors.
- 4) *Solutions to ill-structured problems typically are not right or wrong*, and not valid or invalid....
- 5) *Solution usually is justified by ... **argument** ...* and perhaps ... a rebuttal... attacking a particular constraint or barrier ... or by [refuting] an anticipated opposing position.
- 6) *Evaluation of a solution [is] in terms of agreement or disagreement*, acceptability or the lack thereof, or some level of plausibility.
- 7) The evaluator ... may disagree with the solver’s justification, the extent to which the justification supports the solution offered, or the evaluator may offer a better solution. (Voss, 2006, pp. 305f)

Legal Example of Ill-Structured Problem

California v. Carney, 105 S. Ct. 2066 (1985)

Issue: Legality, under 4th Amendment, US Constitution, of warrantless search of a motor home.

Facts: Police suspected def. Carney of trading drugs for sex in motor home located in a downtown San Diego parking lot. After questioning a boy leaving Carney's motor home, agents entered without a warrant or consent, observed drugs, and arrested Carney.

Conflicting principles:

- *Prevent evidence loss:*
 - Prevent loss of evidence in emergency situation.
- *Privacy:*
 - Constitutional right of privacy and autonomy in ones home.
- *Police efficiency:*
 - Bright line rule that police can apply efficiently.

Deductive Argument Model

Rule

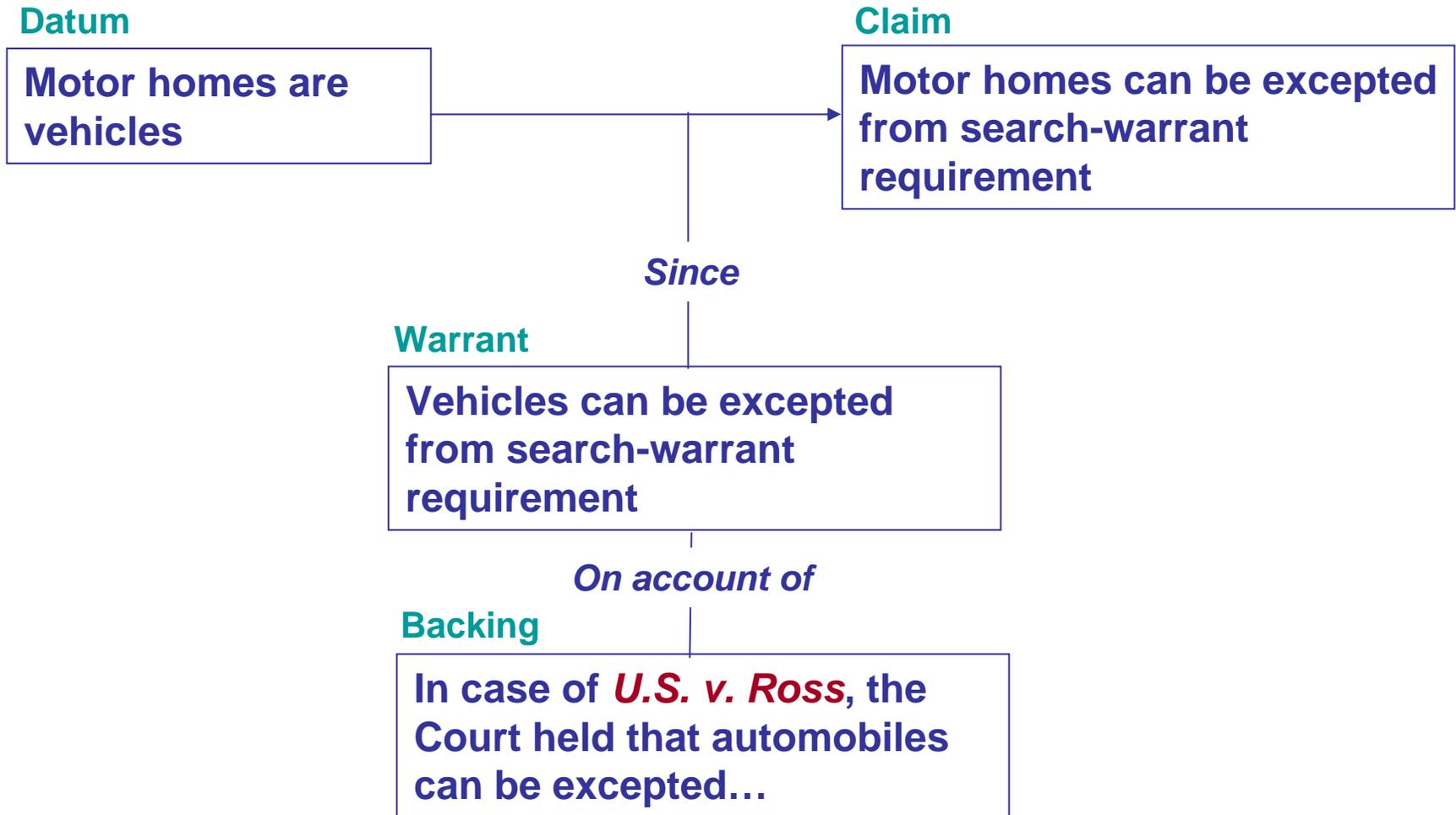
**Vehicles can be excepted
from search-warrant
requirement**

Motor homes are vehicles

Therefore

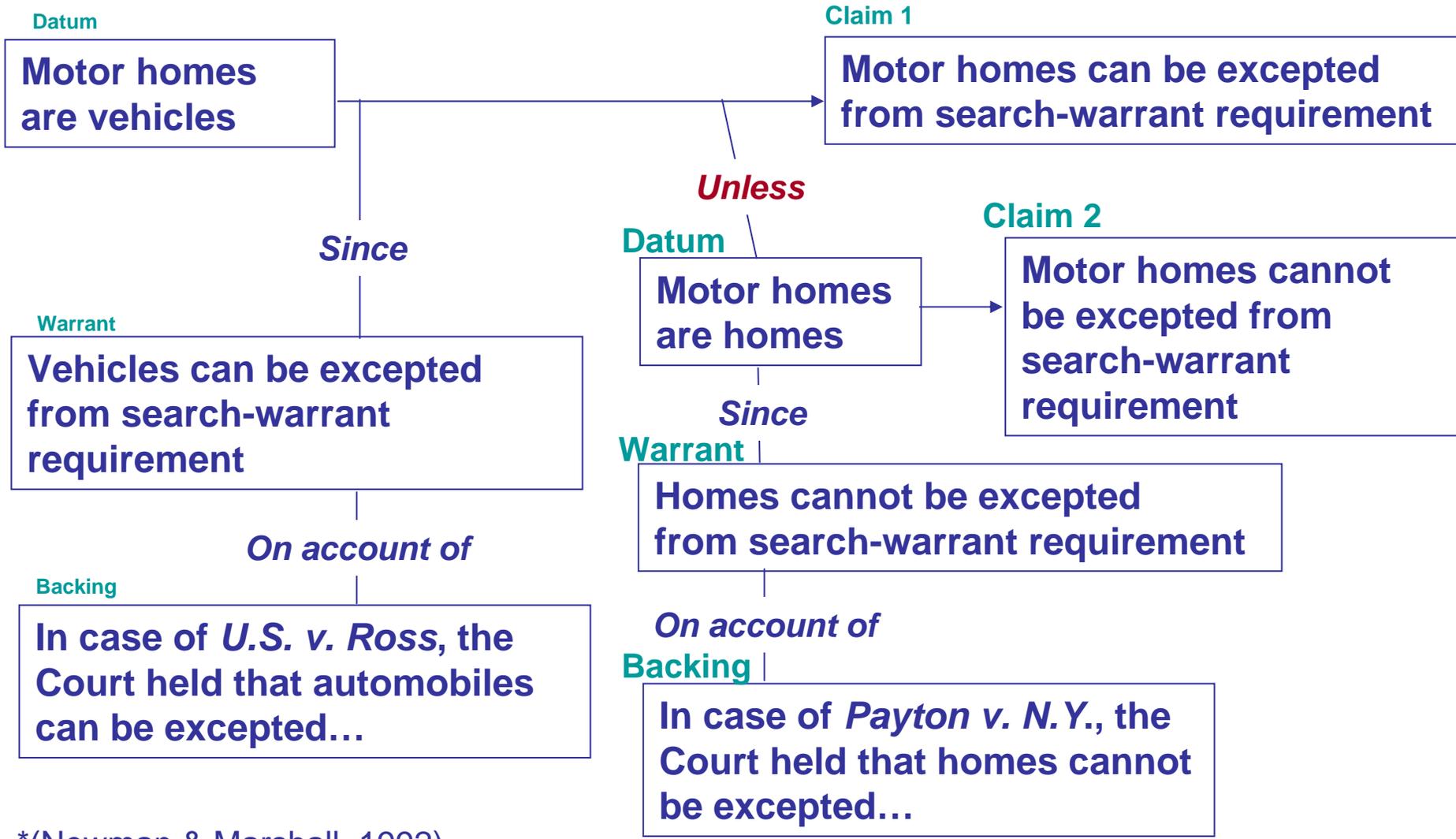
**Motor homes can be excepted
from search-warrant
requirement**

Toulmin Structure for Legal Argument*



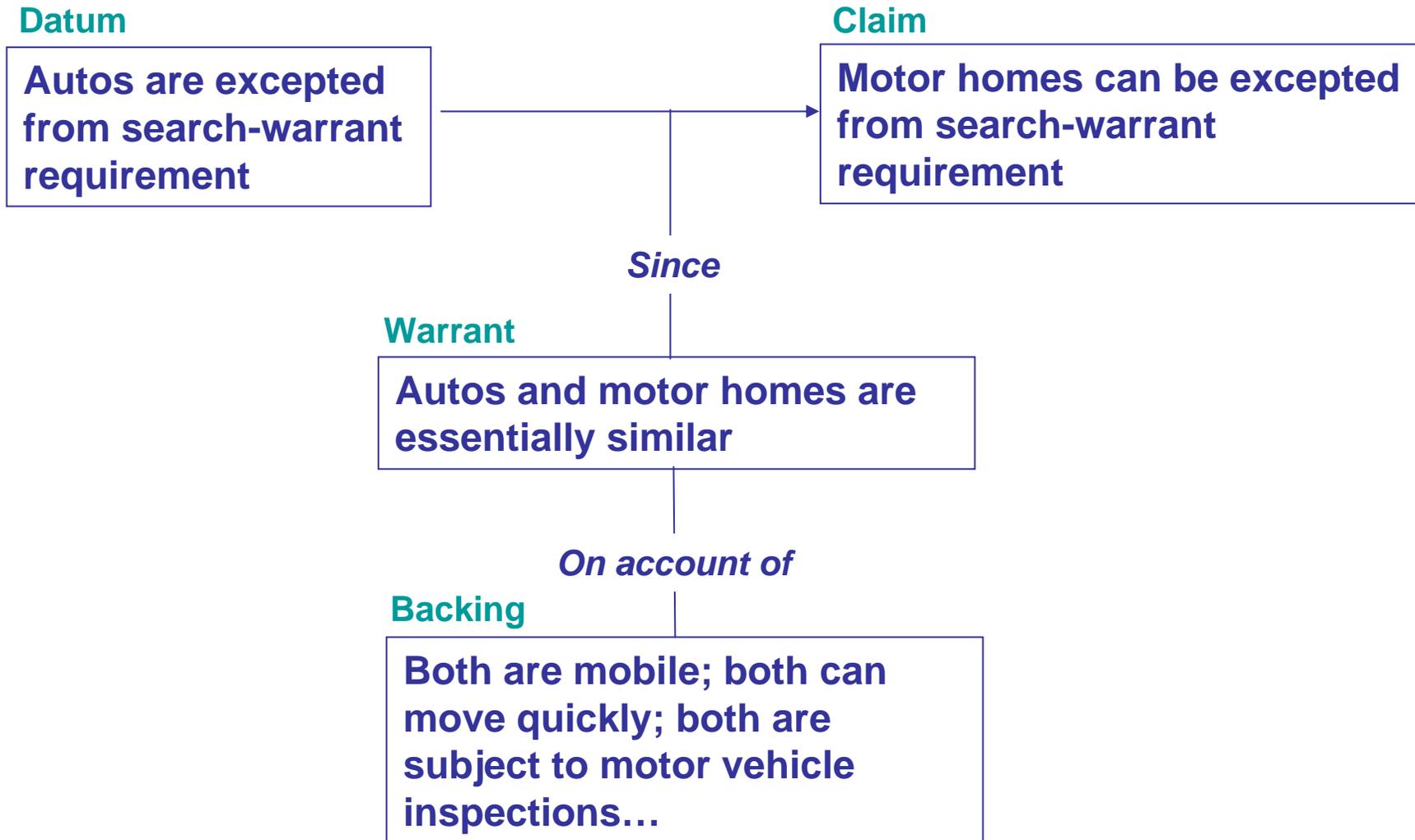
*(Newman & Marshall, 1992)

Toulmin Legal Argument with Rebuttals*



*(Newman & Marshall, 1992)

Toulmin Structure for Analogy*



*(Newman & Marshall, 1992)

Reasoning with Tests and Hypotheticals

Propose test:*

Mr. H: If place-to-search has wheels and is self-propelling → no warrant required.

Carney's motor home has wheels and is self-propelling.

Principle = Prevent evidence loss

Attack test
-- with hypothetical:

J: What if the vehicle is self-propelled but has been in one of these mobile home parks for three months and it's hooked up to water and electricity but still has its wheels on?

Principle = Privacy; Privacy may trump Preventing evidence loss

Abandon test

Modify test

Justify test

-- by analogizing hypo to *Carney*:

Mr. H: society is not willing to recognize that expectation of privacy as justifying a different rule from another motor vehicle; and that, because of its mobility, the capacity for it to move...

Also, both are self-propelled and police cannot know how long the place-to-search has been there, attached to the plumbing, etc.

Principle = Police efficiency (bright-line test); Preventing evidence loss + Police efficiency may trump Privacy

* Proposed test is a **warrant**

What do students need to learn about arguing?

- Not just deductive reasoning with rules...
- Rules (i.e., proposed tests or warrants) have sources; **Argue about:**
 - rules' backing
 - “right” formulation of the rule
- Applying a rule is an interpretive step; **Argue:**
 - if/how rule applies to fact situation
 - about what rule's term means (since they are not adequately defined)
 - Distinguish term's technical vs. common sense meanings
- Applying a rule is a normative conclusion; **Argue:**
 - about policies/principles underlying the regulation
 - how well result “fits” policies/principles, cases and hypotheticals
 - what similarities/differences are relevant
 - how much weight sims/diffs have in light of underlying policies/principles

Why teach argumentation with computer-supported argument diagrams?

Reification: Making argumentation models explicit is assumed to help students understand what they should learn.

Keeping track: Easier for students to track what supports/responds to what.

Practice: Gives students additional opportunities to practice analyzing arguments or engaging in argumentation.

Collaboration: Supports collaborative argumentation among students.

Reflection: Recording argument enables reflection on meaning of argument and how to evaluate it.

Modeling: Provides context for computationally modeling sophisticated reasoning and evaluating the argument models.

Does instruction with computer-supported argument diagrams really work?*

Belvedere: *Trend but not stat/significant:*

- Students using Toulmin-based diagrams of their own scientific arguments produced better post-test arguments than control
- its evidential strength, inferential difficulty, inferential spread, comparison to expert. (Suthers & Hundhausen, 2003)

CSAV: *Trend but not stat/significant:*

- Over semester, students using computer-supported Toulmin-based diagrams of their own legal arguments made higher quality arguments than controls. (Carr, 2003)

Reason!Able, et al.: *Uncontrolled comparison:*

- Diagramming reconstructed arguments from textual examples improves scores on Critical Thinking Skills Tests compared to college or critical thinking courses. (Twardy, 2004; van Gelder, 2007; see Harrell, 2007 [suggesting hand-drawn diagrams work, too])

Convince Me:

- Subjects trained using program to diagram explanatory relations among data and hypotheses made more coherent post-test arguments than controls
- as measured by ECHO. (Schank, 1995)

*(Braak, et al. 2006)

[Dis] Advantages of Toulmin Diagrams

- Capture argument's functional or "propositional" structure
- Go beyond logical deductive inference
- Extendable to case-based, analogical warrants and backings
- Accommodate rebuttals, argument chains, hierarchical argument structure, conjunctive arguments...

BUT

- Quickly generate "spaghetti":
 - too finely grained
 - hard to accommodate recursive structures
- Where do students need more help:
 - relating claims to data? vs. formulating/interpreting warrants?
- Miss important ways to interpret warrants:
 - Lack strategic argument processes (e.g., hypothetical reasoning)
 - Lack dynamic change in interpretation of the warrants

Accommodating Strategic Processes in Toulmin Diagrams

(Voss, et al., 1983) developed graphical coding of expert/novice solutions to ill-structured problems like, “How to solve the problem of Soviet Agriculture.”

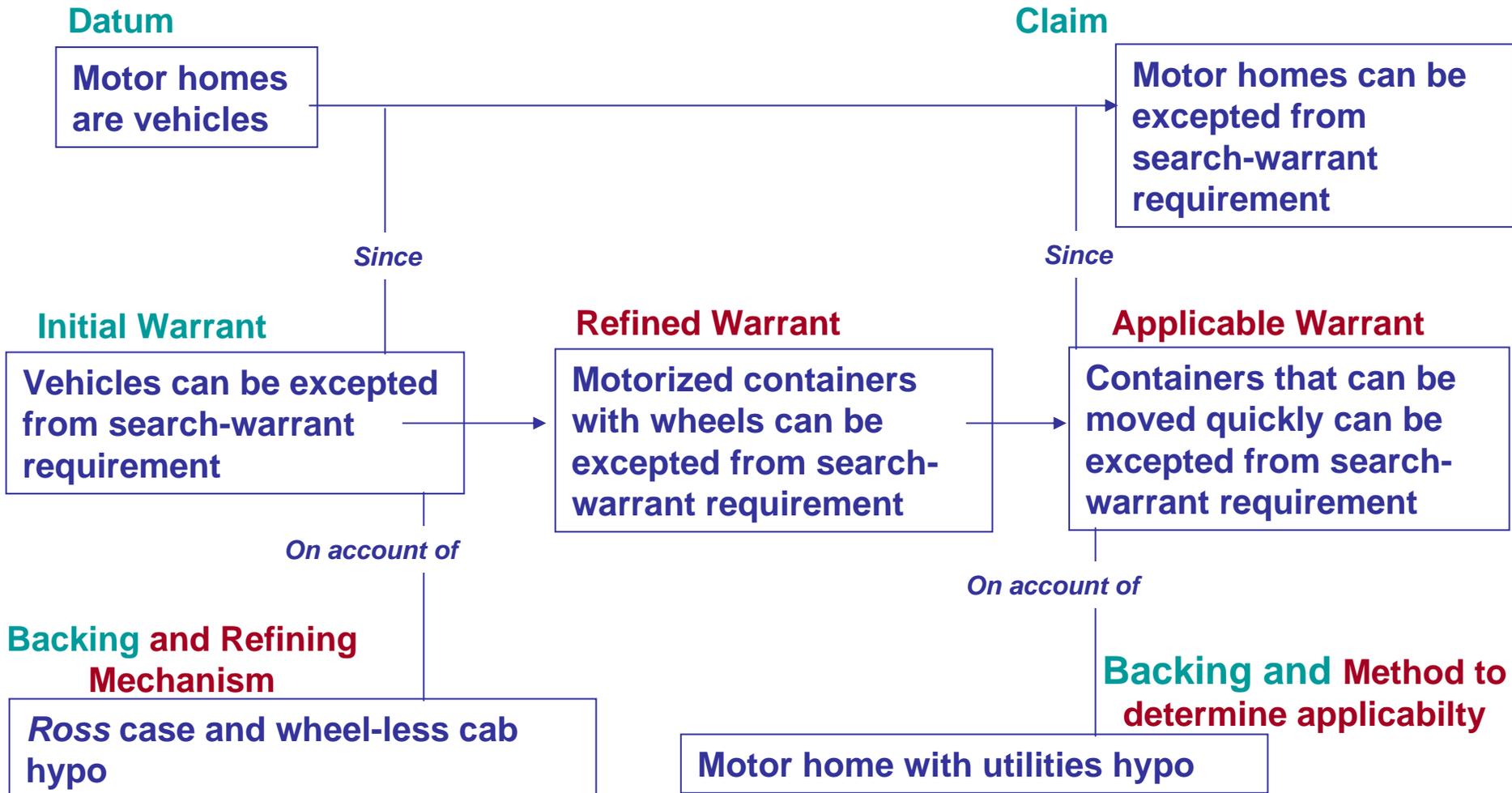
G Structure: **G**eneral problem-solving
Control operators

| | |
|-------------|------------------------------------|
| GCON | State constraint |
| GSUB | State subproblem |
| GSOL | State solution |
| <i>GIPS</i> | <i>Interpret problem statement</i> |
| <i>GSUP</i> | <i>Provide support</i> |
| <i>GEVA</i> | <i>Evaluate</i> |
| <i>GSUM</i> | <i>Summarize</i> |

R Structure: **R**easoning with Toulmin-style
argument operators

| | |
|------|--------------------------|
| RARG | State argument |
| RSAS | State assertion |
| RFAC | State fact |
| RPSC | Present specific case |
| RREA | State reason |
| ROUT | State outcome |
| RCOM | Compare and/or contrast |
| RELA | Elaborate and/or clarify |
| RCON | State conclusion |
| RQUA | State qualifier |

Accommodating Warrant Versions in Toulmin Structure for Legal Arguments*



*(Tans, 2006)

Supplementing Argument Diagrams: Compare Competing Warrants*

Principles / Policies

| | Prevent evidence loss | Protect privacy | Police efficiency: bright-line test | |
|----------------|---|-----------------|--|----|
| Proposed tests | Except entities with wheels | ++ | -- | ++ |
| | Except self-propelling entities | ++ | -- | + |
| | Except places to reside that move quickly | ++ | - | + |
| | Except places to reside that are potentially moveable | + | - | -- |
| | Except self-propelling places to reside subject to DMV inspection | ++ | - | ++ |

*(adapting Newman & Marshall, 1992)

LARGO (Legal ARgument Graph Observer)

- Students reconstruct hypothetical reasoning in SCOTUS oral arguments
- They make argument diagrams:
 - Diagram elements based on a process model of hypothetical reasoning
 - Nodes: Proposed tests, hypotheticals, current facts
 - Links: Relations such as: modified to, distinguished from, analogized to, leads to
- LARGO provides feedback
 - Feedback based on “argument patterns”, text mark-up, and collaborative filtering
 - Detects:
 - important parts of argument text not diagrammed
 - mistaken linkages
 - opportunities for reflection
- Outputs advice prompting students to:
 - Remediate apparently weak parts of diagrams.
 - Reflect on significance of relations among tests, hypotheticals, and responses.

Model for reasoning with hypotheticals

→ **Propose test and argument for deciding case:**

Construct a proposed test that leads to a favorable decision in the case and is consistent with applicable legal principles and important past cases, and give reasons.

← **Pose hypothetical example to probe proposed test:**

Construct a hypothetical example that is analogous to [disanalogous from] the case (i.e., a suitable test when applied to the example should yield the same [a different] result) and yet the proposed test when applied to the example leads to a different [the same] result, and give reasons.

→ **Rebut or otherwise reply to hypothetical example:**

- (1) Save the proposed test by showing that the supposedly disanalogous [analogous] example is really **analogous** [disanalogous]). Or
- (2) **Distinguish example**; **Modify the proposed test** so that it behaves like a suitable test or does not apply to the example. Or
- (3) Abandon the proposed test.

LARGO Approach

Argument transcript

Search : next

an officer.

80. Then secondly, there is the more fundamental problem of which one of these parking spaces is or is not entitled to the added protection.

81. QUESTION: May I inquire, just so I understand your position? Is it that the vehicle have wheels? Could a trailer without a tractor in front of it qualify?

82. MR. HANOIAN: No, I don't think it would, Your Honor, because it would be more or less like the suitcase.

83. QUESTION: I'm sorry? What is your position. You tell me your position.

84. MR. HANOIAN: Our position is that if the officer looks at this conveyance and determines that it has the objective indicia of mobility --

85. QUESTION: [*13] Now does that mean self-propelled?

86. MR. HANOIAN: Self-propelled.

87. QUESTION: It has to be self-propelled?

88. MR. HANOIAN: Yes. I would agree with that.

89. QUESTION: So you wouldn't apply your thought to a trailer park?

90. MR. HANOIAN: Not when it's parked, no. When it's attached, yes, in the same way that one would --

91. QUESTION: But when what about a self-propelled vehicle that's plugged into the plumbing and the

ADVICE

SHOW LIST OF PAST ADVICE

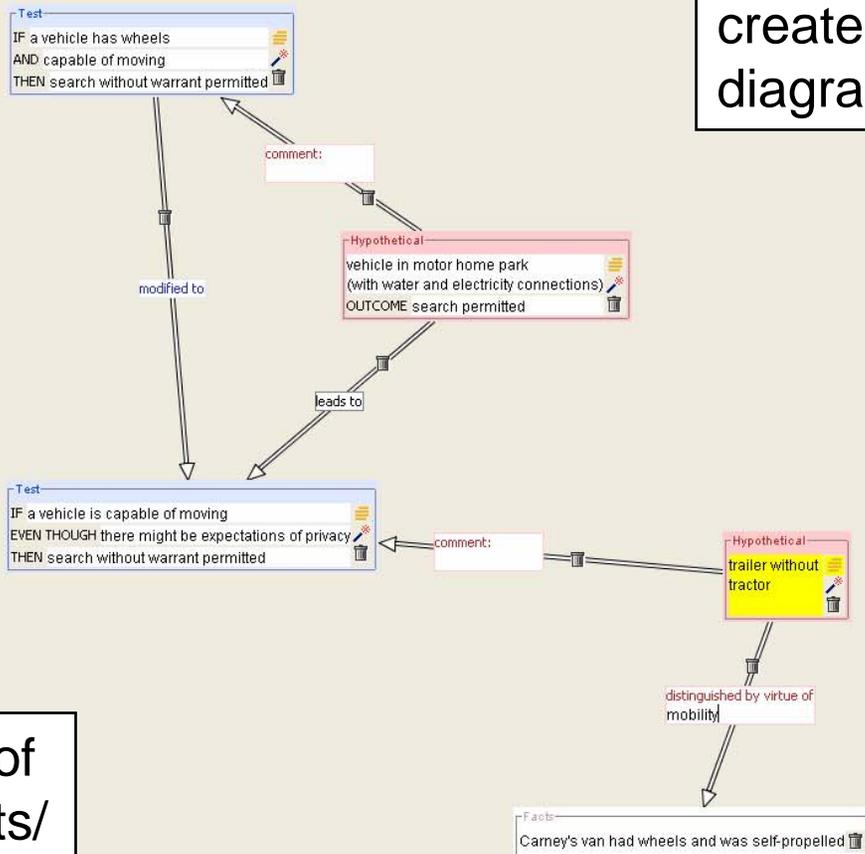
Elements

| | | |
|--------------|------|-------|
| Hypothetical | Test | Facts |
| IF | THEN | |

- Relations
- modified to
 - distinguished from
 - analogized to
 - leads to
 - general relation

Palette of Elements/ Relations

Student-created diagram



LARGO Advice

The screenshot displays the LARGO 1.0 software interface. On the left, a transcript window shows a legal discussion with several paragraphs. Paragraph 66a and 66b are highlighted in yellow. Below the transcript, an 'Advice' dialog box is open, offering three options: 'Include the current fact situation in your diagram', 'Reflect on different tests', and 'Include response to hypothetical in diagram'. The 'Hypothetical' and 'Test' sections at the bottom left show a diagram with a red 'Hypothetical' box and a blue 'Test' box. The main workspace on the right contains a logic diagram with three nodes: a blue 'Test' box at the top, a red 'Hypothetical' box at the bottom right, and another blue 'Test' box at the bottom left. Arrows indicate relationships between these nodes, with a 'comment:' box pointing to the top 'Test' node. The bottom 'Test' node is highlighted in yellow.

start.xml - LARGO 1.0

Datei

Transcript

Search : next

but it is placed upon on blocks --

65. QUESTION: Well, there are places where people can plug into water, and electricity, and do. There are many places, for example, in the state I came from where people go and spend the winter in a mobile home. And you think there would be no expectation of privacy in such circumstances?

66a. MR. HANOIAN: Well, I am not suggesting that there is no expectation of privacy in those circumstances, Your Honor.

66b. MR. HANOIAN: What I am suggesting is that society -- this Court has determined that society is not willing to recognize that expectation of privacy as justifying a different rule from another motor vehicle, and that, because of its mobility, the capacity for it to move [*10] --

67. QUESTION: Well, it just seemed to me that your approach and that offered by the Solicitor General were rather different in that area.

68. MR. HANOIAN: Yes. That is the precise area where our approaches are different.

69. QUESTION: You prefer a single rationale for the exception to the warrant requirement. Namely, you think "mobility" is practically the sole criteria; and the Solicitor

70. Here are some suggestions that could help you with your tasks. Please click on one to see a detailed explanation.

- Include the current fact situation in your diagram
- Reflect on different tests
- Include response to hypothetical in diagram

Relations

- modified to
- distinguished from
- analogized to
- leads to
- general relation

Test

IF a vehicle has wheels
AND capable of moving
THEN search without warrant permitted

comment:

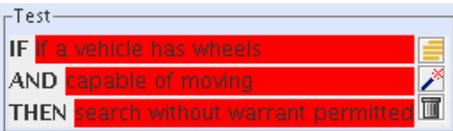
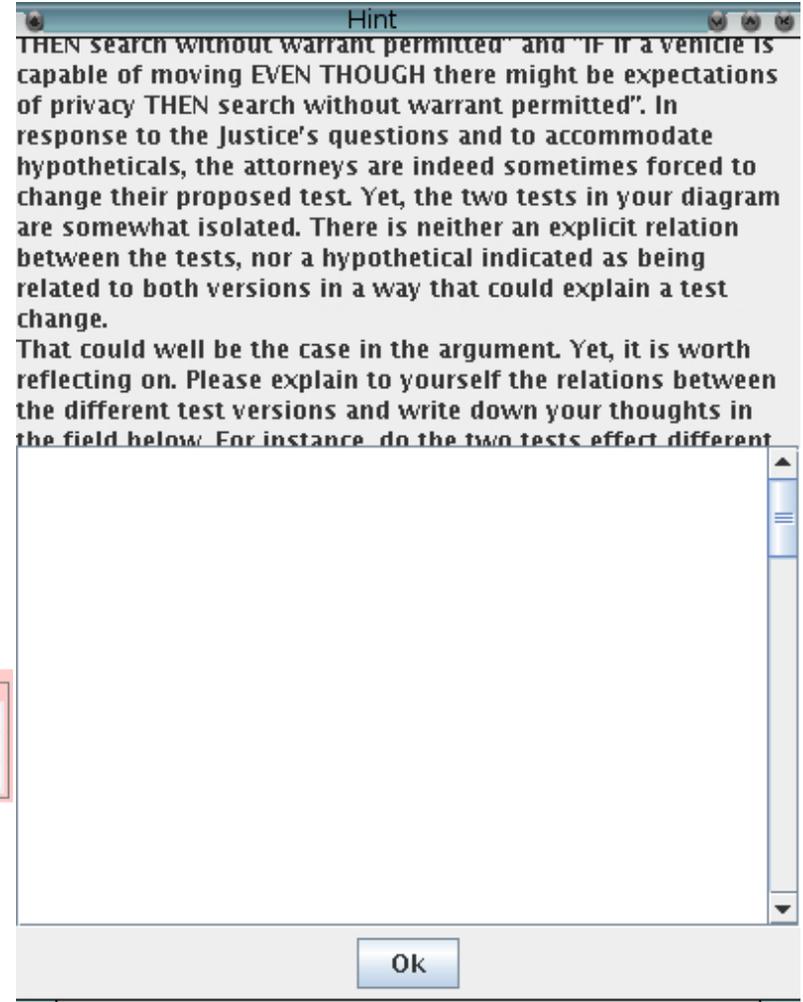
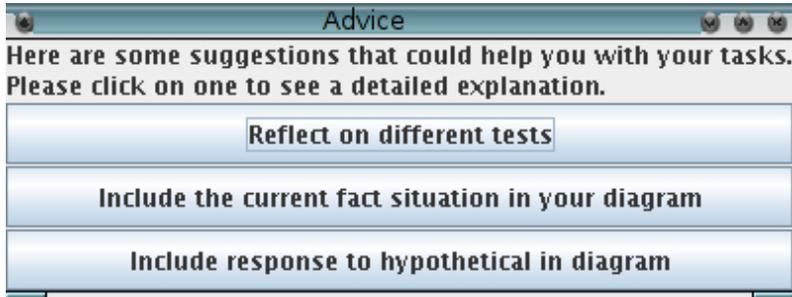
Hypothetical

vehicle in motor home park
(with water and electricity connections)
OUTCOME search permitted

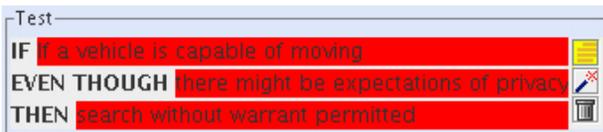
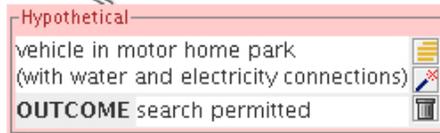
Test

IF a vehicle is capable of moving
EVEN THOUGH there might be expectations of privacy
THEN search without warrant permitted

Advice dialog and self-explanation prompt



comment



Experiments

Goal: Evaluate LARGO as compared to standard note taking.

Hypothesis: LARGO improves understanding of hypothetical reasoning, & ability to recognize/reason about examples in near and far transfer legal domains.

Task: Read SCOTUS oral arguments; represent hypothetical reasoning.

Experimental condition (*Diagram*): Use LARGO graphical argument representation and feedback to id/relate elements of hypothetical reasoning.

Control condition (*Text*): Same oral arguments and focus on hypothetical reasoning, but use text-based word-processing and highlighting.

| | Fall 2006 | Fall 2007 |
|--------------|---|---|
| Participants | 28 1Ls in Legal Process, randomly assigned (38 paid volunteers less 10 failed-to-complete) | 85 1Ls in one Legal Process section, randomly assigned (85 unpaid conscripts less 15 failed to spend time) |
| Procedure | (2h) Pre-test & tool intro w/ <i>Carney</i> example | Same |
| | | (2h) Analyze <i>Asahi</i> case oral arguments (personal jurisdiction) and answer 2 questions |
| | (2h) Analyze <i>Burnham</i> case oral arguments (personal jurisdiction) | Same (+ 2 questions) |
| | (2h) Analyze <i>Burger King</i> case oral arguments (personal jurisdiction) | Same (+ 2 questions) |
| | (2h) Post-test: Near transfer: <i>Keeton</i> case Far transfer: <i>Sony</i> copyright case | Same but no far transfer case. |

Control group: *Text* version of LARGO

The screenshot displays the LARGO 1.0 application window. On the left, a search panel titled 'Transcript' contains a search box and a list of text excerpts. One excerpt is highlighted in yellow: '81. QUESTION: May I inquire, just so I understand your position? Is it that the vehicle have wheels? Could a trailer without a tractor in front of it qualify?'. Below the search panel is a 'Highlight color' section with radio buttons for 'yellow' (selected) and 'white'. The main window area on the right shows a transcript of a legal discussion, with lines 48-50, 60, 66, and 81 visible. The status bar at the bottom indicates 'Page 9 Sec 1 9/9 At 6.1" Ln 4 Col 1' and shows the Windows taskbar with the time 12:54 PM.

Argument transcript

Panel for note taking

Highlighting tools

PLUS:
Instructions about model of hypothetical reasoning and extended example.

Experimental Results

| Fall 2006: 28 paid, volunteer 1Ls in Legal Process class | Fall 2007: 85 unpaid conscripted 1Ls in one Legal Process section |
|---|---|
| <p>Lack of overall result:</p> <p>LARGO did <i>not</i> improve learning across whole sample.</p> <p>Trend favored <i>Diagram</i> but not significant.</p> | <p>No significant differences on post-test between <i>Diagram</i> and <i>Text</i>.</p> <p><i>Text</i> gained significantly more than <i>Diagram</i> on shared, counterbalanced personal jurisdiction questions.</p> |
| <p>Among lower-LSAT students:</p> <p><i>Diagram</i> better than <i>Text</i> on legal issues questions and near-transfer problems re selecting proposed tests, hypotheticals and responses.</p> <p>LOW + MED LSAT <i>Diagram</i> better than <i>Text</i> on evaluating hypotheticals wrt a proposed test.</p> | <p>Not observed.</p> |
| <p>Use of Help:</p> <p>Increase in use of advice function over time</p> <p>High use of advice function (10.1 per case)</p> | <p>Decrease in use of advice function over time.</p> <p>Much lower rate of advice function usage (1.8 per case)</p> |
| <p>Attending to text:</p> <p><i>Diagram</i> students found more important parts of text.</p> | <p><i>Diagram</i> students performed worse than 2006 counterparts.</p> |

Why don't argumentation ITSs work better?

- Representational mismatch:
 - Toulmin-structure vs. diagrams tailored to strategic argument processes for reasoning about warrants
 - Cognitive/pedagogical mismatch of visual diagrams for textual arguments
 - Diagrams reify structure but leave out information
 - Need to interpret visual signs
 - Some arguments are about texts, e.g., proposed legal rules are texts
 - Post-test design:
 - Do diagrams prime students for textual questions better than texts? vs.
 - Can students use diagrams more effectively than text?
- Learning by “observing” argumentation vs. making arguments
 - Ecological validity
 - Law schools teach argumentation skills with oral argument practice
 - Arguing against the program (See e.g., CATO-Dial: Ashley, et al., 2002)
- Motivational problems
 - Argument diagrams vs. rich rhetorical, social context (cf. Willard, 1976)
 - Getting students to use on-demand help for ill-structured problems
 - Volunteer vs. required activity
 - Reconstructing arguments vs. reconstructing an argument for a reason

How to improve argumentation ITSs like LARGO

- Redesign help to engage even students with low motivation.
 - Highlight in advance parts of diagram where LARGO can give help.
 - Encourage competition
 - Report how well student's diagram covers important parts of text as compared with peers' diagrams.
- Engage students more directly in making arguments.
 - Stop action after a Justice's hypothetical:
 - ask **students** to make advocate's best response
 - compare students' responses with advocate's real response.
 - Model arguing with hypotheticals in a parallel domain
 - a simpler, game-like context where a program can pose the hypos.
- Explore **diagnostic uses** of argument diagrams.
 - What can diagrams tell about students' understanding of arguments?
 - Can diagrams be analyzed automatically for diagnostic purposes?

3d Experiment: Are LARGO diagrams diagnostic?

Goal: Compare LARGO diagrams made by 3L students with 1Ls.

Hypothesis: LARGO diagrams of 3Ls reflect more mature understanding of argumentation than 1Ls'.

Procedure: Same as Fall 2007, but with no control condition.

Task: Read SCOTUS oral arguments; represent hypothetical reasoning.

Experimental condition (Diagram): Use LARGO graphical argument representation and feedback to id/relate elements of hypothetical reasoning.

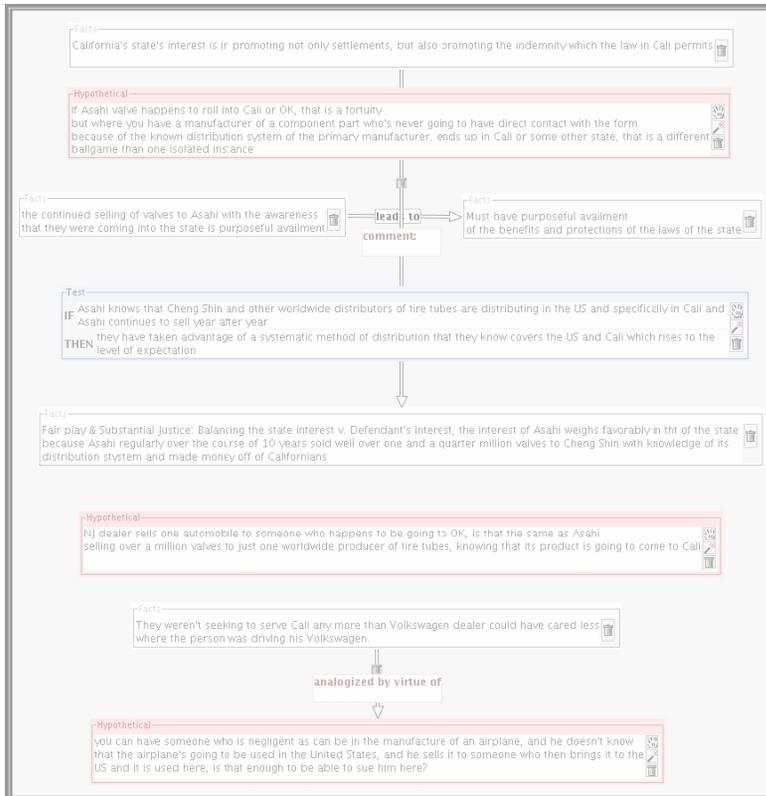
Control condition: Not applicable.

Analysis: Compare diagrams across studies (2006, 2007, 2008) in relation to populations (volunteer 1Ls, nonvolunteer 1Ls, 3Ls), LSATs, and post-test scores.

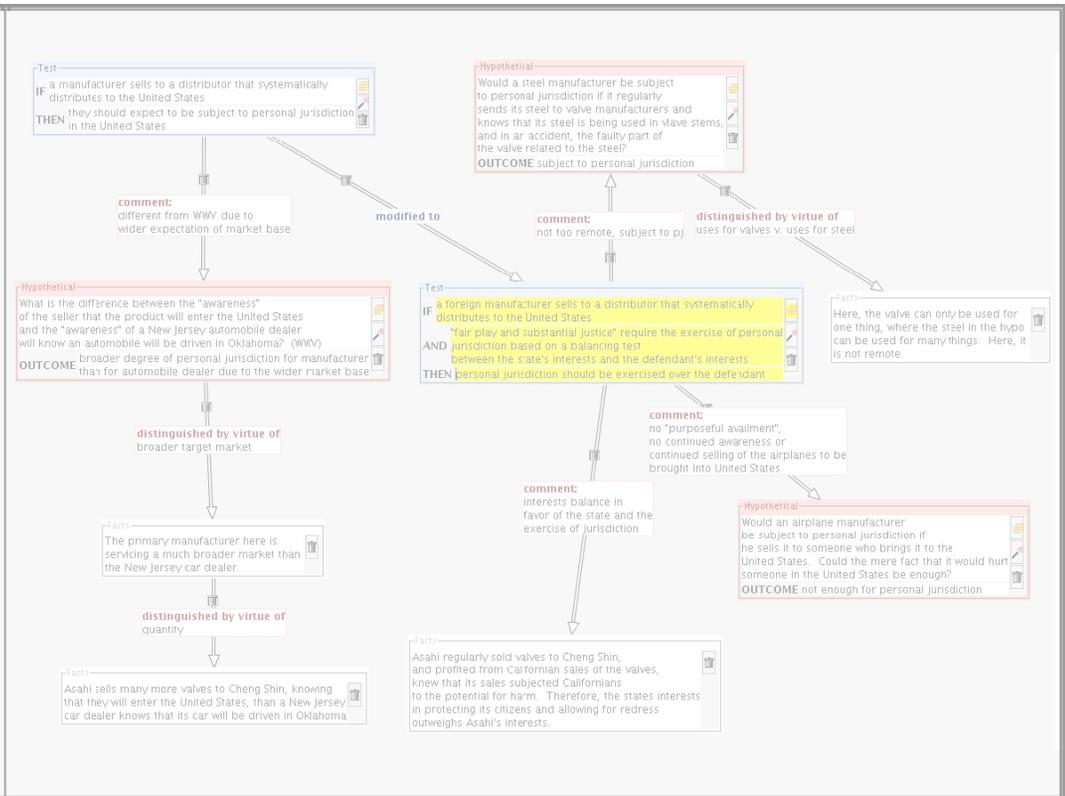
| | |
|---------------------|--|
| Participants | Spring 2008 |
| | 25 3Ls (paid volunteers) |
| Procedure | (2h) Pre-test & tool intro w/ <i>Carney</i> example |
| | (2h) Analyze <i>Asahi</i> case oral arguments (personal jurisdiction) and answer 2 questions |
| | (2h) Analyze <i>Burnham</i> case oral arguments (personal jurisdiction) and answer 2 questions |
| | (2h) Analyze <i>Burger King</i> case oral arguments (personal jurisdiction) and answer 2 questions |
| | (2h) Post-test: Near transfer: <i>Keeton</i> case |

Student diagrams are different!

Are the differences diagnostic?



1st semester 1L student



3L student

How diagnostic are LARGO diagrams?

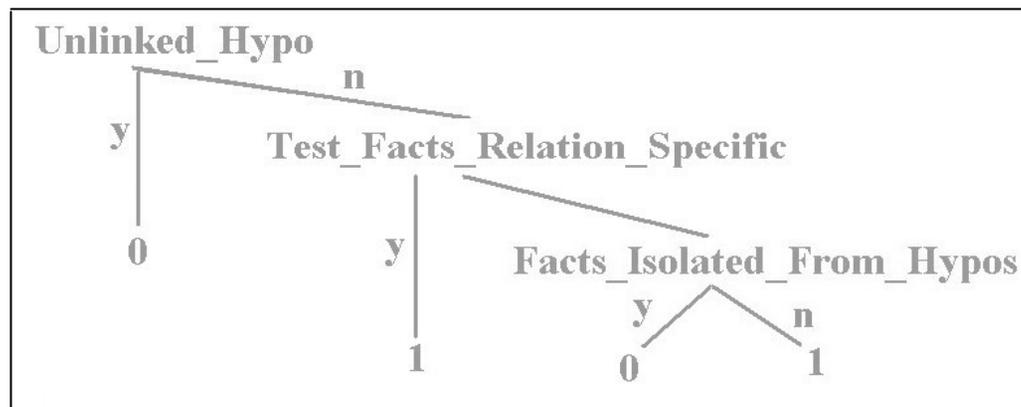
- Diagrams of different abilities/experience differ:
 - **Link-to-node ratio** in '07 correlates positively with LSAT scores ($r=.32, p<.05$).
 - Similar trend in 06-study but not with 3Ls in 08-study.
 - 3Ls' diagrams':
 - **#relations** ($m=12.3$) significantly more ($p<.05$) than volunteer 1Ls' ($m=7.9$)...
 - ... who produce significantly more than non-volunteer 1Ls ($m=5.2$).
 - **#elements** ($m=10.5$) and 1L volunteers' ($m=9.6$) were significantly more ($p<.05$) than 1L non-volunteers ($m=7.5$).
 - **link-to-node ratio** (avg. 1.14) significantly larger ($p<.05$) than 1Ls' diagrams (avg. .82, 67).
- LARGO diagram characteristics that best classify students by:
 - Post-test performance (“above-median” v. “below-median”):
 - **Unlinked-hypo** (Chi-square, $p < 0.01$, precision=47/51)
 - **Unlinked-test** (Chi-square, $p < 0.001$, precision=38/51)
 - **Test-revision-suggested** (Chi-square, $p < 0.01$, precision=35/51)
 - Group membership (3L v. 1L):
 - **No-facts** (Chi-square, $p < 0.01$, precision=32/51)
 - **Unlinked-test** (Chi-square, $p < 0.05$, precision=32/51)
 - **Test-revision-suggested** (Chi-square, $p < 0.001$, precision=41/51)
 - **Test-facts-relation-specific** (Chi-square, $p<0.01$, precision=39/51)

Using ML to classify argument diagrams automatically

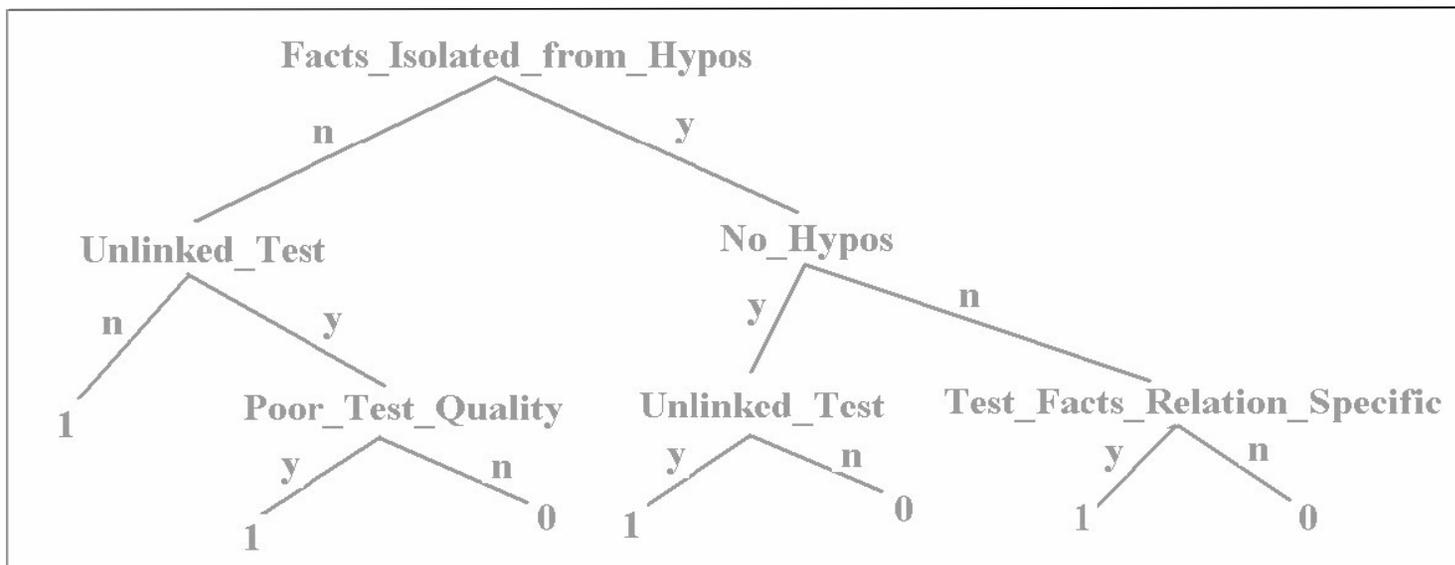
Hypothesis: ML techniques can generate pedagogically useful classification rules to predict posttest performance.

Data: 51 graph/test pairs (34 1Ls, 17 3Ls)

Procedure: Train/test split (90/10): train ML algorithm on portion of data; test on remainder.



C4.5 Median Decision Tree: 86.7% test cases correct



GP Median Decision Tree: 89% test cases correct (best of number of alternative trees)

Why diagnostic diagrams matter

- Misconceptions about argument strategies hard to identify.
 - Instructors miss subtle errors in intermittent oral performance.
- Law professors use arguments to teach *substantive* law.
 - This assumes students understand argumentation.
 - Diagnostic diagrams help check if assumption is true!
 - Don't wait to final exam to find out.
- Automating analysis enables:
 - flagging students who have not understood argumentation.
 - ITSs like LARGO to target help better.
 - identifying new diagnostic patterns.

Continuing Investigation of Diagram's Diagnostic Uses

- Compare early vs. late diagrams
- Analyze relation of help usage to changes in diagrams.
- Apply machine learning genetic algorithm to identify diagnostic rules that predict performance.
- Blinded legal instructors will “grade” diagrams and generate evaluation criteria.

Conclusions

- Argumentation is closely related to reasoning, especially for ill-structured problem solving. Students need to learn argumentation to learn reasoning.
- While intuitively, computer-supported argument diagrams should help teach argumentation, the results have been mixed.
- Toulmin-style argument diagrams may not be ideal representations for arguments about how to formulate and interpret warrants.
- Our LARGO program teaches an argument strategy for reasoning about warrants with diagrams tailored to arguing with hypotheticals.
 - Even so, learning results have still been mixed.
- Especially in ill-structured domains, argumentation ITSs need better techniques for engaging students in using advice and making arguments.
- One emerging role for argument diagrams in instruction is as a diagnostic tool.
 - LARGO diagrams are related to some measures of student ability and success.

Selected References

- Braak, S.W. van den, Oostendorp, H. van, Prakken, H. & Vreeswijk, G. (2006). A critical review of argument visualization tools: Do users become better reasoners?. In F. Grasso, R. Kibble & C. Reed (Eds.), *Workshop Notes of the ECAI-06 Workshop on Computational Models of Natural Argument (CMNA-06)* (pp. 67-75).
- Chad S. Carr, *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*, chapter Using computer supported argument visualization to teach legal argumentation, 75–96, Springer-Verlag, London, UK, 2003.
- Harrell, M. (2007) Using Argument Diagramming Software to Teach Critical Thinking Skills, EISTA (submitted).
- Newman, S. & Marshall, C. (1992) Pushing Toulmin Too Far: Learning From an Argument Representation Scheme. Xerox PARC Tech. Rpt. SSL-92-45.
- Schank, P. (1995). Computational tools for modeling and aiding reasoning: Assessing and applying the Theory of Explanatory Coherence. (Doctoral dissertation, University of California, Berkeley, 1995). Dissertation Abstracts International.
- Daniel D. Suthers and Christopher D. Hundhausen, ‘Learning by constructing collaborative representations: An empirical comparison of three alternatives’, in *European Perspectives on Computer-Supported Collaborative Learning, Proceedings of the First European Conference on Computer-Supported Collaborative Learning*, eds., P. Dillenbourg, A. Eurelings, and K. Hakkarainen, pp. 577–584, Maastricht, the Netherlands, (2001).
- Tans, O. (2006) „The Fluidity of Warrants: Using the Toulmin Model to Analyze Practical Discourse“ in *Arguing on the Toulmin Model: New Essays in Argument Analysis and Evaluation* David Hitchcock and Bart Verheij (ed.) (2006) Springer
- Twardy, C. Argument maps improve critical thinking. *Teaching Philosophy*, 27:95–116, 2004.
- van Gelder, T. The Rationale for Rational™ in Tillers, P. ed., *Law, Probability and Risk* (2007) 6,(1–4):23-42, Special Issue on Graphic and Visual Representations of Evidence and Inference in Legal Settings.
- Voss, J. Toulmin's Model and the Solving of Ill-Structured Problems in *Arguing on the Toulmin Model: New Essays in Argument Analysis and Evaluation* David Hitchcock and Bart Verheij (ed.) (2006) Springer
- Voss, J. F., Greene, T. R., Post, T. A., & Penner, B. C. C. (1983). Problem solving skill in the social sciences. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research theory* (Vol. 17, pp. 165-213). New York: Academic Press.
- Voss, J.F. and M. Means (1991) “Learning to Reason via Instruction in Argumentation”, *Learning and Instruction* 1, 337-350.
- Willard, C.A., (1976) *On the Utility of Descriptive Diagrams For the Analysis and Criticism of Arguments. Communication Monographs*, Vol. 43.

Selected LARGO References

- Ashley, K. (2007) “Interpretive Reasoning with Hypothetical Cases”, Proceedings of The 20th International FLAIRS Conference, Special Track on Case-Based Reasoning, Key West, Florida, USA. May.
- Ashley, K., Pinkwart, N., Lynch, C., & Alevén, V. (2007). Learning by Diagramming Supreme Court Oral Arguments. In *Proceedings of the 11th International Conference on Artificial Intelligence and Law* (p. 271-275). New York (NY), ACM Press.
- Lynch, C., Ashley, K., Pinkwart, N., & Alevén, V. (2007). Argument diagramming as focusing device: does it scaffold reading? In V. Alevén, K. Ashley, C. Lynch, & N. Pinkwart (Eds.), *Proceedings of the Workshop on AIED Applications for Ill-Defined Domains at the 13th International Conference on Artificial Intelligence in Education* (p. 51-60). Los Angeles (CA).
- Pinkwart, N., Alevén, V., Ashley, K., & Lynch, C. (2007). Evaluating Legal Argument Instruction with Graphical Representations Using LARGO. In R. Luckin, K. Koedinger, & J. Greer (Eds.), *Proceedings of the 13th International Conference on Artificial Intelligence in Education* (p. 101-108). Amsterdam (The Netherlands), IOS Press.
- Pinkwart, N., Ashley, K., Alevén, V. and Lynch, C. (2008) “Graph Grammars: an ITS Technology for Diagram Representations” To appear in Proceedings of The 21st International FLAIRS Conference, Special Track on Intelligent Tutoring Systems, Coral Gables, Florida, USA. May.
- Pinkwart, N., Lynch, C., Ashley, K., and Alevén, V. (2008) “Reevaluating LARGO in the Classroom: Are Diagrams Better than Text for Teaching Argumentation Skills?” To appear in Proceedings of the 9th International Conference on Intelligent Tutoring Systems. Montreal, June.