Teaching Argumentation with Computers: One step back, two steps forward?

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 Interaction 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

- AI & Law needs ways to evaluate computational models of argument,
  - e.g., do they really help students learn argumentation skills?
- Computer-supported instruction with argument diagrams
  - has shown only suggestive evidence of success.
  - Argument diagrams that link data and claims through warrants may not be ideal.
- LARGO ITS diagrams focus on argument strategies and processes for reasoning about the warrants.
- Caveat: So far, evidence of learning with LARGO is mixed at best, too!
- Why? Reluctance to use LARGO’s advice
  - Apparently due to motivational issues
  - Yet students’ argument diagrams varied greatly in interesting ways.
  - Systematically compared diagrams of novice vs. more experienced students.
- What did we learn? Argument diagrams provide diagnostic evidence of students’ understanding of argumentation.
  - Could lead to early intervention in law school and better targeting of ITS advice...
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 seem to be missing:
  – [dis]advantages of Toulmin diagrams
• LARGO approach:
  – Teaches arguing with hypotheticals
  – Report of a series of evaluations
• Why don’t LARGO and other 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


**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
Motor homes are vehicles

Vehicles can be excepted from search-warrant requirement

In case of *U.S. v. Ross*, the Court held that automobiles can be excepted...

*(Newman & Marshall, 1992)*
Motor homes are vehicles

Vehicles can be excepted from search-warrant requirement

In case of U.S. v. Ross, the Court held that automobiles can be excepted...

*(Newman & Marshall, 1992)
Toulmin Structure for Analogy*

Datum

Autos are excepted from search-warrant requirement

Claim

Motor homes can be excepted from search-warrant requirement

Since

Warrant

Autos and motor homes are essentially similar

On account of

Backing

Both are mobile; both can move quickly; both are subject to motor vehicle inspections…

*(Newman & Marshall, 1992)
Propose test:*  
Mr. H: If place-to-search has wheels and is self-propelling → no search-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

---

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 terms mean (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 argument components and their relations 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  
– *ito* 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

• May quickly generate “spaghetti”:
  – too fined 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: General problem-solving Control operators**

<table>
<thead>
<tr>
<th>GCON</th>
<th>State constraint</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSUB</td>
<td>State subproblem</td>
</tr>
<tr>
<td>GSOL</td>
<td>State solution</td>
</tr>
<tr>
<td>GIPS</td>
<td>Interpret problem statement</td>
</tr>
<tr>
<td>GSUP</td>
<td>Provide support</td>
</tr>
<tr>
<td>GEVA</td>
<td>Evaluate</td>
</tr>
<tr>
<td>GSUM</td>
<td>Summarize</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>RARG</th>
<th>State argument</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSAS</td>
<td>State assertion</td>
</tr>
<tr>
<td>RFAC</td>
<td>State fact</td>
</tr>
<tr>
<td>RPSC</td>
<td>Present specific case</td>
</tr>
<tr>
<td>RREA</td>
<td>State reason</td>
</tr>
<tr>
<td>ROUT</td>
<td>State outcome</td>
</tr>
<tr>
<td>RCOM</td>
<td>Compare and/or contrast</td>
</tr>
<tr>
<td>RELA</td>
<td>Elaborate and/or clarify</td>
</tr>
<tr>
<td>RCON</td>
<td>State conclusion</td>
</tr>
<tr>
<td>RQUA</td>
<td>State qualifier</td>
</tr>
</tbody>
</table>
Accommodating Warrant Versions in Toulmin Structure for Legal Arguments*

**Datum**
- Motor homes are vehicles

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

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

**Refined Warrant**
- Motorized containers with wheels can be excepted from search-warrant requirement

**Applicable Warrant**
- Containers that can be moved quickly can be excepted from search-warrant requirement

**Backing and Refining Mechanism**
- Ross case and wheel-less cab hypo

**On account of**

**Since**

*(Tans, 2006)*
## Supplementing Argument Diagrams: Compare Competing Warrants*

### Principles / Policies

<table>
<thead>
<tr>
<th>Proposed tests</th>
<th>Prevent evidence loss</th>
<th>Protect privacy</th>
<th>Police efficiency: bright-line test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Except entities with wheels</td>
<td>++</td>
<td>- -</td>
<td>++</td>
</tr>
<tr>
<td>Except self-propelling entities</td>
<td>++</td>
<td>- -</td>
<td>+</td>
</tr>
<tr>
<td>Except places to reside that move quickly</td>
<td>++</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Except places to reside that are potentially moveable</td>
<td>+</td>
<td>-</td>
<td>- -</td>
</tr>
<tr>
<td>Except self-propelling places to reside subject to DMV inspection</td>
<td>++</td>
<td>-</td>
<td>++</td>
</tr>
</tbody>
</table>

*(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:
  – Remedy apparently weak parts of diagrams.
  – Reflect on significance of relations among tests, hypotheticals, and responses.
Process Model of Hypothetical Argument-Pt.1

→ 1. Propose test: For proponent, propose test for deciding the current fact situation (cfs):
   Construct a proposed test that leads to a favorable decision in the cfs and is consistent with applicable underlying legal principles/policies and important past cases, and give reasons.

← 2. Pose hypothetical: For interlocutor, pose hypothetical example to probe if proposed test is too broad:
   Construct a hypothetical example that:
   (a) emphasizes some normatively relevant aspect of the cfs and
   (b) to which the proposed test applies and assigns the same result as to the cfs, but
   (c) where, given legal principles/policies, that result is normatively wrong in the hypothetical.

→ 3. Respond: For proponent, respond to interlocutor’s hypothetical example showing test too broad:
   (3.a) Justify the proposed test: Analogize the hypothetical example and the cfs and argue that they both should have the result assigned by the proposed test. Or
   (3.b) Modify the proposed test: Distinguish the hypothetical example from the cfs, argue that they should have different results and that the proposed test yields the right result in the cfs, and add a condition or limit a concept definition so that the narrowed test still applies to the cfs but does not apply to, or leads to a different result for, the hypothetical example. Or
   (3.c) Abandon the proposed test and return to (1) (i.e., construct a different proposed test that leads to a favorable decision in the cfs and is consistent with applicable underlying legal principles/policies, important past cases, and hypotheticals…)
LARGO Approach

Argument transcript

Student-created diagram

Palette of Elements/Relations
LARGO Advice
Advice dialog and self-explanation prompt

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

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

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

THEN search without warrant permitted

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

In response to the justice's questions and to accommodate hypotheticals, the attorneys are indeed sometimes forced to change their proposed test. Yet, the two tests in your diagram are somewhat isolated. There is neither an explicit relation between the tests, nor a hypothetical indicated as being related to both versions in a way that could explain a test change.

That could well be the case in the argument. Yet, it is worth reflecting on. Please explain to yourself the relations between the different test versions and write down your thoughts in the field below. For instance, do the two tests effect different
LARGO Advice Characteristics

• Elements or Relations Not Used
  – NO ELEMENTS (1)
  – NO TESTS (1)
  – NO HYPOS (1)
  – NO FACTS (1)
  – NO RELATIONS (2)

• Elements Not Linked to Argument Text
  – UNLINKED HYPO (2)
  – UNLINKED TEST (2)

• Element Descriptions Not Filled-in
  – LINKED TEST WITHOUT TEXT (2)
  – LINKED HYPO WITHOUT TEXT (2)
  – TEST WITHOUT CONDITION (2)
  – TEST WITHOUT CONCLUSION (2)
  – DISTINGUISH WITHOUT TEXT (2)
  – ANALOGIZE WITHOUT TEXT (2)

• Known Elements Missed in Text
  – MISSED TEST (2)
  – MISSED HYPO (2)

• Mislabeled Elements in Text
  – HYPO AS TEST (4)
  – TEST AS HYPO (4)
  – IRRELEVANT AS TEST (4)
  – IRRELEVANT AS HYPO (4)

• Make Relations More Specific
  – HYPO RELATION GENERAL (3)
  – HYPO FACTS RELATION GENERAL (3)
  – HYPO MULTIPLE LEADSTO (4)

• Mistaken Relations
  – TEST RELATION NOT MODIFY (3)
  – TEST HYPO RELATION MODIFIED (4)
  – HYPO TEST RELATION MODIFIED (4)
  – TEST FACTS RELATION SPECIFIC (4)
  – FACTS TESTS RELATION SPECIFIC (4)
  – CYCLIC REFERENCES (4)
  – HYPO FACTS RELATION MODIFIED (4)
  – FACTS RELATIONS (4)
  – HYPO FACTS RELATION CAUSES (4)

• Possible Missing Relations
  – AMBIGUOUS HYPO RELATION (3)
  – TEST ISOLATED FROM HYPOS (3)
  – ISOLATED HYPO (3)
  – FACTS ISOLATED FROM HYPOS (3)
  – TEST FACTS WITHOUT HYPOS (3)
  – TESTS MODIFIED WITHOUT REASON (3)
  – NO RESPONSE TO HYPO (3)
  – UNRELATED TESTS (5)
  – HYPO HYPO RELATION NO RESPONSE (5)

• Consider Significance of Elements & Relations
  – DISCUSS HYPO ONE TEST (5)
  – DISCUSS HYPO MULTIPLE TESTS (5)
  – ISOLATED HYPO DISCUSS (5)
  – MULTIPLE IMPACT OF HYPO (5)
  – TEST TRIGGERS COLLAB FILTERING (5)
  – POOR TEST QUALITY (5)
**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.

<table>
<thead>
<tr>
<th>Participants</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fall 2006</strong></td>
<td>28 1Ls in Legal Process, randomly assigned (38 <strong>paid volunteers</strong> less 10 failed-to-complete)</td>
</tr>
<tr>
<td></td>
<td>(2h) Pre-test &amp; tool intro w/ Carney example</td>
</tr>
<tr>
<td></td>
<td>(2h) Analyze <em>Burnham</em> case oral arguments (personal jurisdiction)</td>
</tr>
<tr>
<td></td>
<td>(2h) Post-test: Near transfer: <em>Keeton</em> case Far transfer: <em>Sony</em> copyright case</td>
</tr>
<tr>
<td><strong>Fall 2007</strong></td>
<td>70 1Ls in one Legal Process section, randomly assigned (85 <strong>unpaid conscripts</strong> less 15 failed to spend time)</td>
</tr>
<tr>
<td></td>
<td>Same</td>
</tr>
<tr>
<td></td>
<td>(2h) Analyze <em>Asahi</em> case oral arguments (personal jurisdiction) and answer 2 questions</td>
</tr>
<tr>
<td></td>
<td>Same (+ 2 questions)</td>
</tr>
<tr>
<td></td>
<td>Same (+ 2 questions)</td>
</tr>
<tr>
<td></td>
<td>Same but no far transfer case.</td>
</tr>
</tbody>
</table>
Control group: *Text version of LARGO*

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

Argument transcript

Panel for note taking

Highlighting tools
## Experimental Results

<table>
<thead>
<tr>
<th>Fall 2006: 28 paid, <strong>volunteer</strong> 1Ls in Legal Process class</th>
<th>Fall 2007: 70 unpaid <strong>conscripted</strong> 1Ls in one Legal Process section</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lack of overall result:</strong></td>
<td>• No significant differences on post-test between <em>Diagram</em> and <em>Text</em>.</td>
</tr>
<tr>
<td>• LARGO did <em>not</em> improve learning across whole sample.</td>
<td>• <em>Text</em> gained significantly more than <em>Diagram</em> on shared, counterbalanced personal jurisdiction questions.</td>
</tr>
<tr>
<td>• Trend favored <em>Diagram</em> but not significant.</td>
<td>• Not observed.</td>
</tr>
<tr>
<td><strong>Among lower-LSAT students:</strong></td>
<td></td>
</tr>
<tr>
<td>• <em>Diagram</em> better than <em>Text</em> on legal issues questions and near-transfer problems re selecting proposed tests, hypotheticals and responses.</td>
<td>• Decrease in use of advice function over time.</td>
</tr>
<tr>
<td>• LOW + MED LSAT <em>Diagram</em> better than <em>Text</em> on evaluating hypotheticals wrt a proposed test.</td>
<td>• Much lower rate of advice function usage (1.8 per case)</td>
</tr>
<tr>
<td><strong>Use of Help:</strong></td>
<td>• No correlation between help use and performance</td>
</tr>
<tr>
<td>• Increase in use of advice function over time</td>
<td></td>
</tr>
<tr>
<td>• High use of advice function (10.1 per case)</td>
<td></td>
</tr>
<tr>
<td>• Help use correlated with learning.</td>
<td></td>
</tr>
<tr>
<td><strong>Attending to text:</strong></td>
<td>• <em>Diagram</em> students performed worse than 2006 counterparts.</td>
</tr>
<tr>
<td>• <em>Diagram</em> students found more important parts of text than <em>Text</em>.</td>
<td></td>
</tr>
</tbody>
</table>
Why don’t argumentation ITSs work better?

Re LARGO:

• **Motivational problems** getting students to use on-demand help for ill-structured problems
  - Volunteer vs. required activity
  - Reconstructing arguments vs. reconstructing an argument for a reason

Generally:

• **Representational mismatch:**
  - Toulmin-structures vs. diagrams for process of reasoning about warrants
  - Argument diagrams vs. rich rhetorical, social context (cf. Willard, 1976)
  - Cognitive/pedagogical mismatch of 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 some one, even a program (See e.g., CATO-Dial: Ashley, et al., 2002)
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 diagrams of student vs. peers cover important parts of text.

• 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**

| 17 3Ls (paid volunteers) (25 – 8 not ready at time of analysis) |

**Procedure**

1. **Pre-test & tool intro w/ Carney example**
2. **Analyze Asahi case oral arguments (personal jurisdiction) and answer 2 questions**
3. **Analyze Burnham case oral arguments (personal jurisdiction) and answer 2 questions**
4. **Analyze Burger King case oral arguments (personal jurisdiction) and answer 2 questions**
5. **Post-test:** Near transfer: *Keeton* case
Student diagrams are different!
Are the differences diagnostic?

1st semester 1L student

3L student
How diagnostic are LARGO diagrams?

- Diagrams of students of different abilities/experience differ...
- In terms of some quantitative measures:
  - **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. 0.82, 67).
- **LARGO advice characteristics** that best classify students by:
  - **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)
  - **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)
Classifying 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 Mean Decision Tree:** 82.2% test and 100% training cases classified correctly.

---

**Genetic Programming Mean Decision Tree:** 87% test cases and 100% training cases classified correctly.
Why diagnostic diagrams matter

• 3L vs. 1L diagram differences relate to “thinking like a lawyer”
  – E.g., attention to text, analogizing/distinguishing, formulating tests, comparing facts
  – When do students learn to “think like a lawyer”?
  – Zone of proximal development? 10/23/2006 study vs. 9/5/2007 study

• 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; insist on compliance
  – identifying new diagnostic patterns.
How can the learning of argumentation skills be assessed objectively?

• Scores on critical thinking tests
• Quality of constructed arguments as assessed by blinded experts
• Quantity/quality of discussion
• Coherence of arguments
• Comparisons of expert/novice arguments
• Special purpose multiple choice tests focusing on argument
• Analysis of arguments for transfer problems
• Analysis of argument diagrams
Continuing Investigation of Diagram’s Diagnostic Uses

- Compare earlier vs. later 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 to “grade” diagrams and generate evaluation criteria:
  1. Can experts grade diagrams as better/worse and what are their criteria?
  2. Can experts diagnose specific issues wrt students’ understanding based on diagrams (e.g., relation of facts to hypotheticals, role of tests)?
Case-based legal reasoning models

• Rules may be:
  – derived from cases and applied deductively, but participants argue about what the rules mean;
  – challenged, changed, and reinterpreted through processes of case comparison.

• Case as an authoritative source of a:
  – rule vs.
  – result given its facts from which to extract a range of rules in light of precedents and underlying principles.

• Test is:
  – proposed for the cfs that deductively leads to a desired decision.
  – subjected to challenge with hypothetical examples that tease out its meaning and assess its fit with the past decisions and principles.
  – applied deductively to the hypothetical and prior case facts, but that is only part of the process.
Conclusions

• Especially in ill-structured problem solving, students need to learn argumentation to learn reasoning.
• While 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.
• Argumentation ITSs need better techniques for engaging students in using advice and making arguments.
• Emerging role for argument diagrams as a diagnostic tool.
  – LARGO diagrams are related to some measures of student ability and success.
  – Could help tailor instruction to student’s zone of proximal development.
• Evaluation methodology could be applied to other argument diagram approaches.
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).


Selected LARGO References


Hypothetical Reasoning in Civil Law Context

• Civil law scholars discuss decisions in treatises.
  – Critique bad decisions to guide judges in future decision-making.
  – Pose situations with undesirable consequences resulting from decision.
  – “In performing their roles as organizers, rationalizers and critics of precedent, academics in some systems in the study make extensive use of hypothetical cases in their work…Indeed, it is a major technique used in the United Kingdom and in the United States, and also in most civil law countries…. D. MacCormick and R. Summers (ed.) Interpreting Precedents, pp. 528-9, Ashgate/Dartmouth (1997).

• Prior to decision, civil law judges engage in a series of deliberative meetings and talk about precedents.
  – European Court of Justice: Advocate General reviews case law and delivers a treatise reporting an alternate path from problem to solution.
  – French Court of Cassation: Reporting judge drafts alternate judgments leading to different results and submits before the oral hearing.