LARGO, an ITS for Teaching Argumentation with Hypotheticals

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 Universität Clausthal

Research sponsored by NSF Award IIS-0412830
What do first year law students need to learn about legal argument?

- 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
- SCOTUS oral arguments in cases involving course material from the first year curriculum provide some good teaching examples...
LARGO: (Legal ARgument Graph Observer)

- Students reconstruct hypothetical reasoning in SCOTUS oral arguments using...

- *Argument diagrams:*
  - Diagram elements based on a process model of hypothetical reasoning
  - Nodes: Proposed tests, hypotheticals, current facts
  - Links: Relations among elements, for example:
    - Hypothetical leads to a test
    - Test is modified to another version of the test
    - Hypothetical is distinguished from (or analogized to) the current facts

- LARGO provides feedback
  - based on “argument patterns”, text mark-up, and collaborative filtering.
  - Detects:
    - important parts of argument text not diagrammed
    - mistaken or missing relations among elements
    - opportunities for reflection
  - Outputs advice prompting students to:
    - Remediate apparently weak parts of diagrams.
    - Reflect on significance of relations among tests, hypotheticals, and responses.
Why is hypothetical reasoning important?

- Makes arguer’s assumptions explicit.
- Systematic methodology for creative, exploratory reasoning
  - Tools to explore meaning of legal concepts
  - Explores linkages among facts, rules, theory, principles, evaluation criteria
  - Shows how modifying rules (e.g., more and less general in certain relevant ways) affects tradeoffs of conflicting principles
- Modifying/improving hypotheses is a natural outcome.
- Real world methodology, sometimes predictive of real outcomes.
What's hard in learning hypothetical reasoning skills?

• Getting good examples (e.g., Sup. Ct. oral arguments)
• Understanding the oral argument examples
  - Oblique dialogues
    • assume familiarity with case background.
    • ako short hand to avoid explicit references to principles, etc.
  
  - Hard to see/explain interpretive relations in the argument texts:
    1) Identify and formulate proposed tests
    2) Explain how a hypothetical relates to a test and why
    3) Explain how an advocate responds to a hypothetical and why
    4) Evaluate response to hypothetical vis a vis the test

• Making interpretive arguments
  - Hard to invent factual hypo for purpose of testing proposed test.
  - Hard to integrate:
    • Facts
    • Principled reasons why suitable test should reach certain outcome
    • Criteria for making and evaluating arguments and selecting appropriate responses.
Example SCOTUS Argument


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.
Toulmin Legal Argument with Rebuttals*

**Datum**
- **Motor homes are vehicles**

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

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

**Claim 1**

**Unless**
- **Motor homes are homes**

**Datum**
- **Homes cannot be excepted from search-warrant requirement**

**Since**
- **Motor homes cannot be excepted from search-warrant requirement**

**Claim 2**

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

**On account of**
- **In case of Payton v. N.Y., the Court held that homes cannot be excepted...**

*(Newman & Marshall, 1992)*
Reasoning with Tests and Hypotheticals

Propose test:*

Mr. H: If place-to-search has wheels and is self-propelling \(\rightarrow\) 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

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
Process model for reasoning with hypotheticals

→ 1. Advocate: 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. interlocutor: Pose hypothetical example to probe if proposed test is too broad [or too narrow]:
   (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 that result is normatively wrong in the hypothetical.

→ 3. Advocate: Respond to interlocutor’s hypothetical example:
   (3.a) Save 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

Advice
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

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

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

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

Spring 2008

Participants

25 3Ls (paid volunteers)

(2h) Pre-test & tool intro w/ Carney example

(2h) Analyze Asahi case oral arguments (personal jurisdiction) and answer 2 questions

(2h) Analyze Burnham case oral arguments (personal jurisdiction) and answer 2 questions

(2h) Analyze Burger King case oral arguments (personal jurisdiction) and answer 2 questions

(2h) 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 different abilities/experience differ ito 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. .82, 67).

- LARGO advice 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)

Key:
0 = below median
1= above median
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)
  – DISTINGUISHED 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)
Why diagnostic diagrams matter

• 3L vs. 1L diagram differences relate to “thinking like a lawyer”
  - E.g., attention to text, analogizing/distinguishing, formulating tests
  - 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.
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 will “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)?
Conclusions

• Law students need to learn how to formulate and interpret warrants (i.e., proposed tests or rules for deciding a case).
• Our LARGO program teaches an argument strategy for reasoning about warrants with diagrams tailored to arguing with hypotheticals.
• We have some evidence that LARGO helps students with lower LSAT scores learn, but LARGO needs better techniques for engaging students in using advice and making arguments.
• One emerging role for LARGO’s argument diagrams in instruction is 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.
Learning 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>Fall 2006</th>
<th>Fall 2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>28 1Ls in Legal Process, randomly assigned (38 paid volunteers less 10 failed-to-complete)</td>
<td>85 1Ls in one Legal Process section, randomly assigned (85 unpaid conscripts less 15 failed to spend time)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2h) Pre-test &amp; tool intro w/ Carney example</td>
</tr>
<tr>
<td>(2h) Analyze Asahi case oral arguments (personal jurisdiction) and answer 2 questions</td>
</tr>
<tr>
<td>(2h) Analyze Burger King case oral arguments (personal jurisdiction)</td>
</tr>
<tr>
<td>(2h) Post-test: Near transfer: Keeton case Far transfer: Sony copyright case</td>
</tr>
</tbody>
</table>
Learning Experiments

Control group: Text version of LARGO

PLUS: Instructions about model of hypothetical reasoning and extended example.
### Learning Experiments 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><strong>Lack of overall result:</strong></td>
</tr>
<tr>
<td>• LARGO did <em>not</em> improve learning across whole sample.</td>
<td>• No significant differences on post-test between <em>Diagram</em> and <em>Text</em>.</td>
</tr>
<tr>
<td>• Trend favored <em>Diagram</em> but not significant.</td>
<td>• <em>Text</em> gained significantly more than <em>Diagram</em> on shared, counterbalanced personal jurisdiction questions.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Among lower-LSAT students:</strong></th>
<th><strong>Among lower-LSAT students:</strong></th>
</tr>
</thead>
<tbody>
<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>• Not observed.</td>
</tr>
<tr>
<td>• LOW + MED LSAT <em>Diagram</em> better than <em>Text</em> on evaluating hypotheticals wrt a proposed test.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Use of Help:</strong></th>
<th><strong>Use of Help:</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Increase in use of advice function over time</td>
<td>• Decrease in use of advice function over time.</td>
</tr>
<tr>
<td>• High use of advice function (10.1 per case)</td>
<td>• Much lower rate of advice function usage (1.8 per case)</td>
</tr>
<tr>
<td></td>
<td>• Advice usage positively correlated w/ post-test scores</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Attending to text:</strong></th>
<th><strong>Attending to text:</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>• <em>Diagram</em> students found more important parts of text than <em>Text.</em></td>
<td>• <em>Diagram</em> students performed worse than 2006 counterparts.</td>
</tr>
</tbody>
</table>
Learning Experiments: Why doesn't LARGO work better?

- Motivational problems getting students to use on-demand help
  - Volunteer vs. required activity
  - Outside zone of proximal development
  - Reconstructing arguments vs. reconstructing an argument for a reason

- Representational mismatch:
  - 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)
Learning Experiments: How to improve 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?
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


