

---

# Toward Assessing Law Students' Argument Diagrams

## **Collin Lynch**

Graduate Student  
Intelligent Systems Program  
University of Pittsburgh  
Pittsburgh, Pennsylvania, USA  
collinl@cs.pitt.edu

## **Kevin D. Ashley**

Professor of Law and Intelligent  
Systems  
Senior Scientist, Learning Research  
and Development Center  
University of Pittsburgh  
ashley@pitt.edu

## **Niels Pinkwart**

Assistant Professor  
Department of Informatics  
Clausthal University of Technology  
Clausthal, Germany  
np@tu-clausthal.de

## **Vincent Alevan**

Assistant Professor  
Human-Computer Interaction Institute  
Carnegie Mellon University  
Pittsburgh, Pennsylvania, USA  
alevan@cs.cmu.edu

# Why interrater reliability matters re argument diagrams

- Legal argument diagrams meant to accomplish real work in the world:
  - Communicate structure of legal argument in a form that humans understand and with which programs can reason. E.g.,
    - Legal professionals, students, ordinary citizens to construct argument diagrams in debate
    - Students to learn by reconstructing arguments of expert judges or philosophers
- Humans need to understand and evaluate argument diagrams reliably.
  - Can they? It's an empirical question.
- Argument formalisms cannot readily account for all possible variations.
  - Content ambiguity: no single formal specification of what may be contained in a Toulmin warrant.
  - Functional ambiguity: possible to construct logically equivalent arguments in many ways.
- Can't ignore variability when constructing or validating argument diagrams.
  - Complicates process of assessing or comparing pairs of diagrams.
  - Especially when dealing with arguments by students or non-experts.

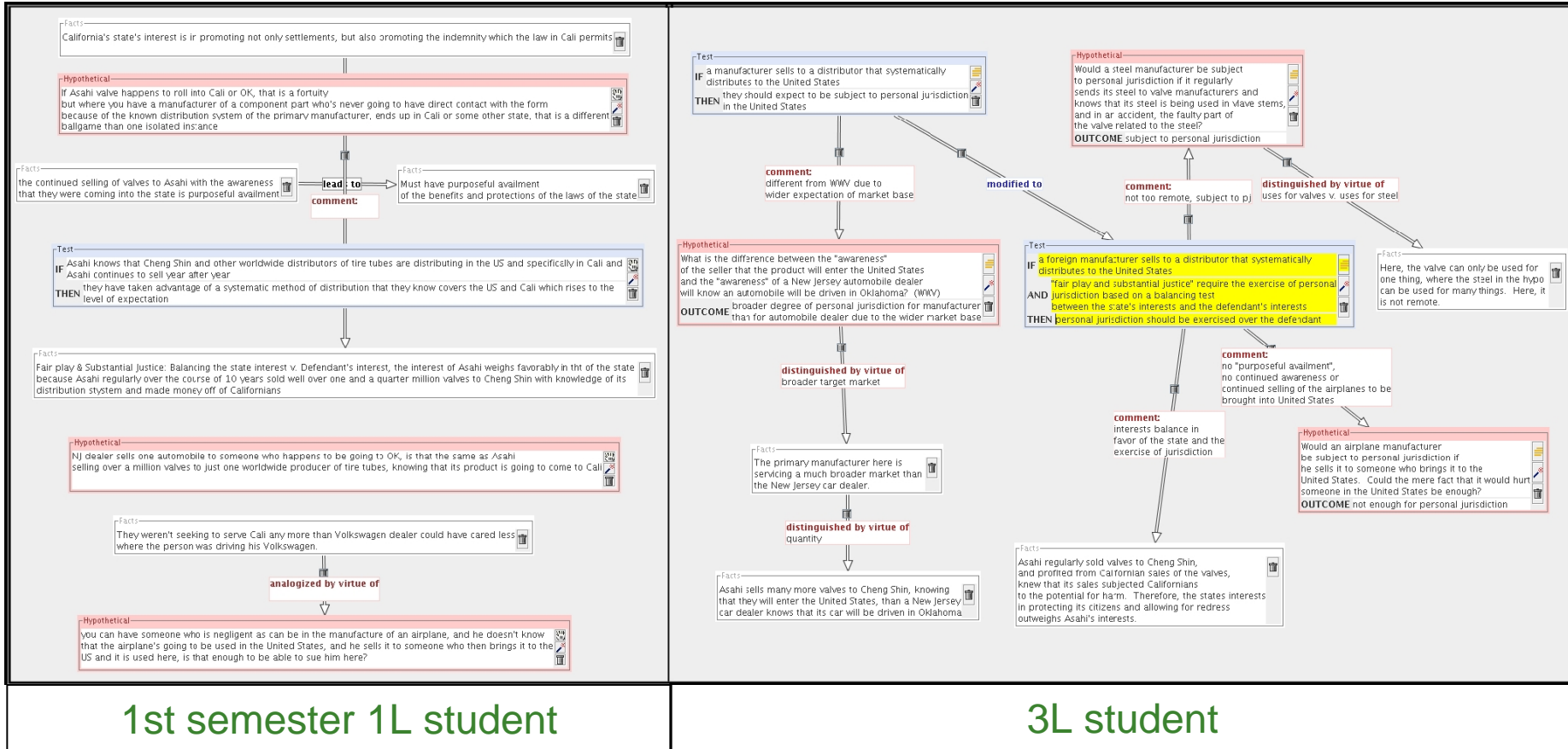
# Experimental Procedure

- In blind test, two law school instructors graded 198 students' argument diagrams of hypothetical reasoning in SCOTUS oral arguments using LARGO.
  - 15 first year (1L) student volunteers from Fall 2006 study in first semester Legal Process course
  - 33 1L non-volunteers from Fall 2007 in one section of Legal Process.
  - 23 third year (3L) student diagrams from middle of 2007-2008 (all paid volunteers upper ranked by GPA)
  - Asahi (57 diagrams), Burger King (71 diagrams), Burnham (71 diagrams)
- Both instructors:
  - trained on LARGO and made own diagrams
  - engaged in practice grading to learn/adjust criteria
- For each case diagram set:
  1. Initial "gestalt" comparison
    - Partition diagrams into poor, medium, good bins.
    - Divide each bin into better and worse
  2. Detailed comparison
    - Reshuffle diagrams
    - Assign detailed grades according to 3 criteria (coverage, correctness, comprehension)
    - Score each test and hypothetical node
  3. Assign **overall grade** on 12 point scale

# Student diagrams are different!

## Are the differences diagnostic?

### Two Student Diagrams of *Asahi* Oral Argument



# Grading Criteria

Category	Criterion
<b>Coverage</b>	How well does the diagram cover ...
	1. ... all of the essential tests in the argument?
	2. ... all of the essential hypotheticals in the argument?
	3. ... all of the essential relationships in the argument?
	4. How well are the diagram components related to the appropriate facts of the case?
	5. ... the argument components as a whole?
<b>Correctness</b>	How well does the diagram...
	1. ... reflect the ways in which the hypotheticals challenge the tests?
	2. ... reflect the ways in which tests are modified in response to hypotheticals?
	3. ... reflect analogizing and distinguishing of hypotheticals with respect to other hypotheticals and essential case facts?
	4. ... capture the role of policies and principles in the argument (e.g., in analogizing and distinguishing)?
	5. Overall, how correctly does the diagram represent the argument?
<b>Comprehension</b>	How well does the student understand...
	1. ... this particular argument both in factual and procedural terms?
	2. ... the role of proposed tests in legal argument?
	3. ... the role of hypothetical cases in argument?
	4. ... the process of analogizing and distinguishing hypothetical cases?
	5. ... the general process of arguing with tests and hypotheticals?
	6. ... the role of policies and principles in arguments of this type?

# Grading Criteria

Category	Criterion
<b>Test Element</b>	1. Is the test summary test like (formulated as a logical rule with applicable conditions and a relevant legal conclusion for deciding an issue or the case)?
	2. Is the test linked to an appropriate segment of the argument?
	3. Is this test correctly related to the relevant preceding tests?
	4. Is this test correctly related to the relevant hypotheticals?
	5. How well does the diagram capture the role this test plays in the argument?
<b>Hypothetical Element</b>	1. How well does the summary reflect the hypothetical posed in the text?
	2. Is this hypothetical correctly related to the relevant test nodes?
	3. How well does the diagram capture the role of this hypothetical in the argument with respect to challenging the tests? For instance, does it capture the judge's implication with the hypothetical (i.e., probing the test as too broad, too narrow, or exploring what the test means)?
	4. How well does the diagram capture the analogizing and distinguishing of this hypothetical with respect to other hypotheticals and essential case facts?

# Preliminary Results

---

- “One way to estimate the reliability of a measure is to compute the correlation between parallel measurements.” Reliability and validity assessment. E. Carmines and R. Zeller (1979) Reliability and Validity Assessment. P. 34 Sage. Thousand Oaks, CA.
- To be completed.

---

# Conclusions

- If argument diagrams are to do real work, humans have to be able to assess them reliably.
- Law students used LARGO to diagram SCOTUS oral arguments involving hypothetical reasoning.
- Instructors grading argument diagrams in blind test for coverage, correctness, and comprehension did so with reasonable reliability.
- Plan to use instructor grades to assess correlations among automatically detected features of diagrams and other measures:
  - LSAT scores, number of years in law school, posttest performance.
- If successful, LARGO diagrams could be tool for early identification of students who do not understand some basics of legal argumentation.