

---

# Connecting Hypothesis-testing to Legal Texts

---

Kevin D. Ashley

Professor of Law and Intelligent Systems

University of Pittsburgh

[ashley@pitt.edu](mailto:ashley@pitt.edu)

# Overview

---

- Ways to connect hypothesis-testing to legal texts.
  - Three kinds of hypotheses lawyers test:
    - Predictive, normative, and semantic
  - Some characteristics of legal hypothesis-testing
- Connecting predictive hypotheses to textual legal cases: SMILE+IBP
  - IBP's issue- and factor-based prediction model
  - SMILE's classifying case texts by factors
- Developing approaches to connect normative or semantic hypothesis-testing to texts.

# Some hypotheses lawyers test

---

1. On the assumed facts, plaintiff should win ...
  - the issue of whether the info is a trade secret
  - in a claim for trade secret misappropriation.
2. In cases of this type, the decision rule should be...
  - “If the thing to be searched has wheels and is self-propelled, then a search warrant is required.”
3. In the documents I’m about to disclose to my opponent, no documents are ...
  - communications between Alice and her lawyer Bob between 1985 and 1989.

# Types of hypotheses lawyers test

---

## 1. Predictive

- On these facts, plaintiff should win ...

## 2. Normative

- In cases of this type, the decision rule should be...

## 3. Semantic

- In the documents I'm about to disclose to my opponent, no documents are...

# Characteristics of lawyers' hypothesis-testing

---

- Need to assess hypotheses in terms of examples and counterexamples.
  - Examples / counterexamples often expressed in text.
  - Result of reasoning with examples / counterexamples often involves modifying the (textual) hypothesis.
- Not just retrieving information but problem-solving with the information retrieved.

# Modeling legal hypothesis-testing

---

## 1. Predictive

- SMILE+IBP
  - IBP: predict outcome based on factors
  - SMILE: classify textual case descriptions by factors

## 2. Normative

- LARGO
  - Model posing hypothetical examples / counterexamples to assess proposed decision rule
  - Annotate argument texts via diagramming

## 3. Semantic

- NSF CISE proposal
  - Induce (dynamic) social network based on document header info
  - Evaluate attorney-constructed hypotheses in terms of social network

# Modeling Predictive Hypotheses: IBP

[Brüninghaus & Ashley ICAIL-03]

*Input:* Current fact situation's factors



Identify issues

Determine favored party for each issue:

- If factors favor same side, return side, else
- Scientific, evidential reasoning with cases:
  - If cases found with issue-related factors
    - Test hypothesis that majority side should win
    - Explain-away counterexamples
  - Otherwise, Broaden-Query

Combine analysis from issues



*Output:* Predicted outcome and explanation

# IBP Domain Model

## Uniform Trade Secrets Act, Restatement of Torts

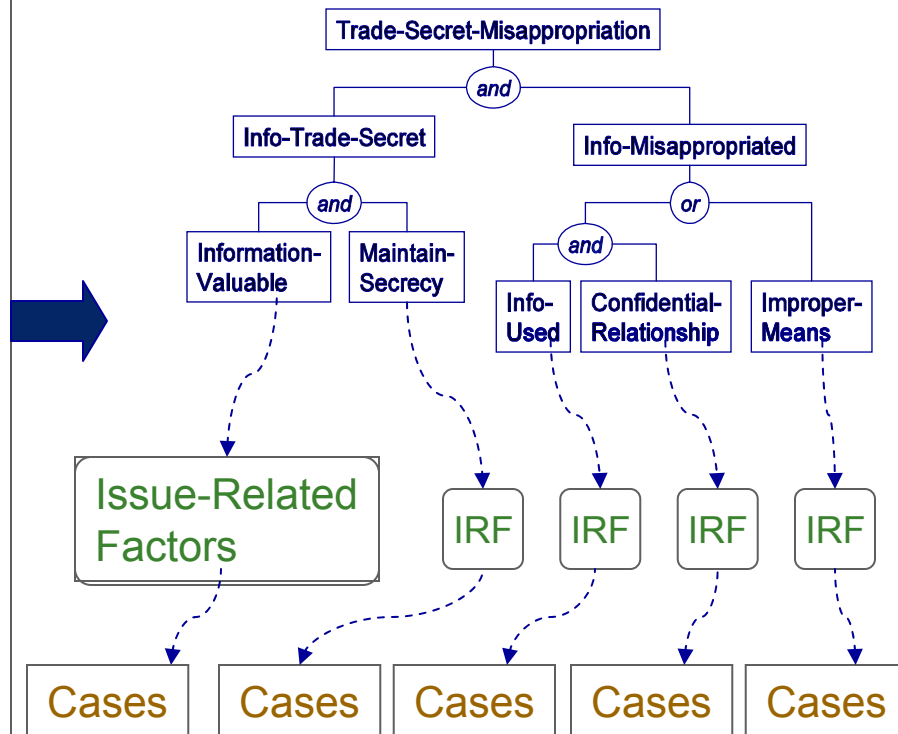
“Trade secret” means information, [...] that:

- (i) derives independent economic value, [...] from not being generally known to, and not being readily ascertainable by proper means [...] and
- (ii) is the subject of efforts that are reasonable under the circumstances to maintain its secrecy.

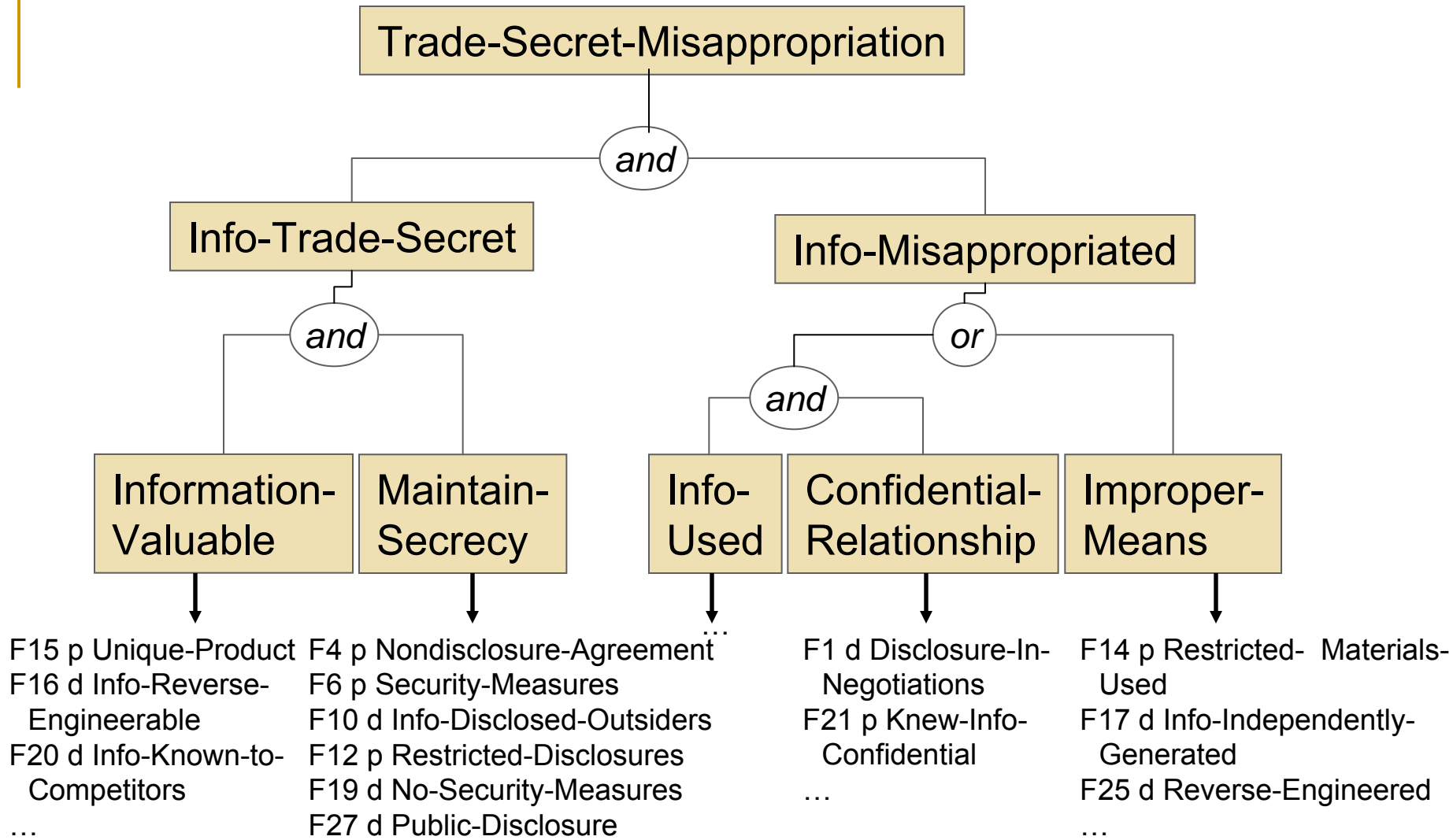
One [...] is liable for trade secret misappropriation if

- (a) he discovered the secret by improper means, or
- (b) his disclosure or use constitutes a breach of confidence [...]

## Logical Structure of Trade Secrets Law



# IBP Domain Model (detail)



# Annotated *National Rejectors* Case Squib (Excerpts)

---

Since the 1940's, National was practically the sole supplier of coin-handling devices, which are used in vending machines, amusement machines, and coin-operated washing machines. [F15] National developed its products (rejectors and changers) through “many years of trial and error, cut and try and experimentation.” In 1957, National employees including defendant Trieman, a sales manager, and Melvin, an engineer, started their own business for producing coin-handling devices. ... Melvin, working at his home, designed two rejectors that were as close as possible to the comparable National rejectors. [F18] ... He also used some National production drawings, as well as a few parts and materials obtained, without consent, from National. [F7] However, none of defendants' drawings was shown to be a copy of a drawing of National. The resulting rejector improved on the National product in certain ways. [Melvin and Trieman resign from National.] National's vice-president testified that the National rejectors could be taken apart simply and the parts measured by a skilled mechanic who could make drawings from which a skilled modelmaker could produce a handmade prototype. [F16] The shapes and forms of the parts, as well as their positions and relationships, were all publicized in National's patents as well as in catalogs and brochures and service and repair manuals distributed to National's customers and the trade generally.[F27] National did not take any steps at its plant to keep secret and confidential the information claimed as trade secrets. [F19] It did not require its personnel to sign agreements not to compete with National. [F19] It did not tell its employees that anything about National's marketed products was regarded as secret or confidential. [F19] Engineering drawings were sent to customers and prospective bidders without limitations on their use. [F10] ...

# IBP Analysis for *National Rejectors* Case

As input: by Human	by SMILE
<p><b>1. Prediction for NATIONAL-REJECTORS</b> Factors favoring plaintiff: (F18 F15 F7) Factors favoring defendant: (F27 F19 F16 F10)</p> <p><b>2. Issue raised in this case is INFO-USED</b> Relevant factors in case: F18(P) F7(P) The issue-related factors all favor the outcome PLAINTIFF.</p> <p><b>3. Issue raised in this case is SECURITY-MEASURES</b> Relevant factors in case: F19(D) F10(D) The issue-related factors all favor the outcome DEFENDANT.</p> <p><b>4. Issue raised in this case is INFO-VALUABLE</b> Relevant factors in case: F27(D) F16(D) F15(P) Theory testing did not retrieve any cases, broadening the query. For INFO-VALUABLE, the query can be broadened for DEFENDANT. Each of the pro-D Factors (F16 F27) is dropped for new theory testing. Theory testing with Factors {F16 F15} gets the following cases: [8 cases won by plaintiff] In this broadened query, PLAINTIFF is favored. Theory testing with Factors {F27 F15} gets the following cases: (DYNAMICS DEFENDANT F4 F5 F6 F15 F27) In this broadened query, DEFENDANT is favored. There is no resolution for INFO-VALUABLE, even when broadening the query.</p> <p><b>5. Outcome of the issue-based analysis:</b> For issue INFO-VALUABLE, ABSTAIN is favored. For issue SECURITY-MEASURES, DEFENDANT is favored. For issue INFO-USED, PLAINTIFF is favored. =&gt; Predicted outcome for NATIONAL-REJECTORS is DEFENDANT</p>	<p><b>1'. Prediction for NATIONAL-REJECTORS</b> Factors favoring plaintiff: (F18 F7 F6) Factors favoring defendant: (F25 F19 F16 F10)</p> <p><b>2'. Issue raised in this case is INFO-USED</b> Relevant factors in case: F25(D) F18(P) F7(P) Theory testing did not retrieve any cases, broadening the query. For INFO-USED, the query can be broadened for PLAINTIFF. Each of the pro-P Factors (F7 F18) is dropped for new theory testing. Theory testing with Factors {F7 F25} still does not retrieve any cases. Theory testing with Factors {F18 F25} gets the following cases: (KG PLAINTIFF F6 F14 F15 F16 F18 F21 F25) (MINERAL-DEPOSITS PLAINTIFF F1 F16 F18 F25) In this broadened query, PLAINTIFF is favored. By a-fortiori argument, PLAINTIFF is favored for INFO-USED.</p> <p><b>3'. Issue raised in this case is SECURITY-MEASURES</b> Relevant factors in case: F19(D) F10(D) F6(P) Theory testing did not retrieve any cases, broadening the query. For SECURITY-MEASURES, query can be broadened for DEFENDANT. Each of the pro-D Factors (F10 F19) is dropped for new theory testing</p> <p>Theory testing with Factors {F10 F6} gets the following cases: <i>[11 cases won by plaintiff, 2 cases won by defendant]</i> Trying to explain away the exceptions favoring DEFENDANT MBL can be explained away with unshared ko-factor(s) (F20). CMI can be explained away with unshared ko-factor(s) (F27 F20 F17). Therefore, PLAINTIFF is favored for the issue.</p> <p>In this broadened query, PLAINTIFF is favored. Theory testing with Factors {F19 F6} still does not retrieve any cases. There is no resolution for SECURITY-MEASURES, even when broadening the query.</p> <p><b>4'. Issue raised in this case is INFO-VALUABLE</b> Relevant factors in case: F16(D) The case has only one weak factor related to the issue, which is not sufficient evidence to include this issue in the prediction.</p> <p><b>5'. Outcome of the issue-based analysis:</b> For issue INFO-USED, PLAINTIFF is favored. For issue SECURITY-MEASURES, ABSTAIN is favored. =&gt; Predicted outcome for NATIONAL-REJECTORS is ABSTAIN</p>

# Focus on IBP's explaining away counterexamples

---

## 3'. Issue raised in this case is SECURITY-MEASURES

Relevant factors in case: F19(D) F10(D) F6(P)

Theory testing did not retrieve any cases, broadening the query.

For SECURITY-MEASURES, query can be broadened for  
DEFENDANT.

Each of the pro-D Factors (F10 F19) is dropped for new theory testing.

Theory testing with Factors {F10 F6} gets the following cases:

*[11 cases won by plaintiff, 2 cases won by defendant]*

Trying to explain away the exceptions favoring DEFENDANT

MBL can be explained away with unshared ko-factor(s) (F20).

CMI can be explained away with unshared ko-factor(s) (F27 F20 F17).

Therefore, PLAINTIFF is favored for the issue.

In this broadened query, PLAINTIFF is favored.

Theory testing with Factors {F19 F6} still does not retrieve any cases.

There is no resolution for SECURITY-MEASURES, even when  
broadening the query.

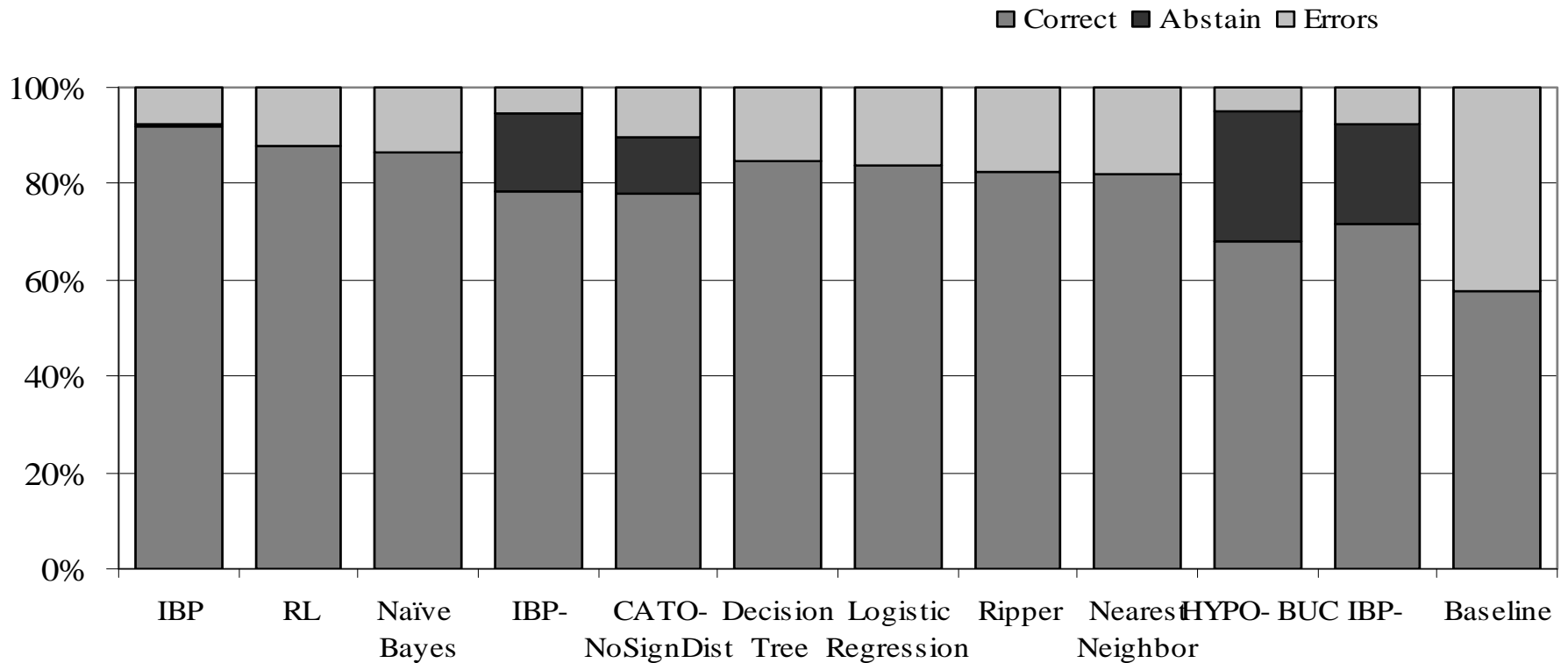
---

# Definition of Knock-Out (KO) Factors

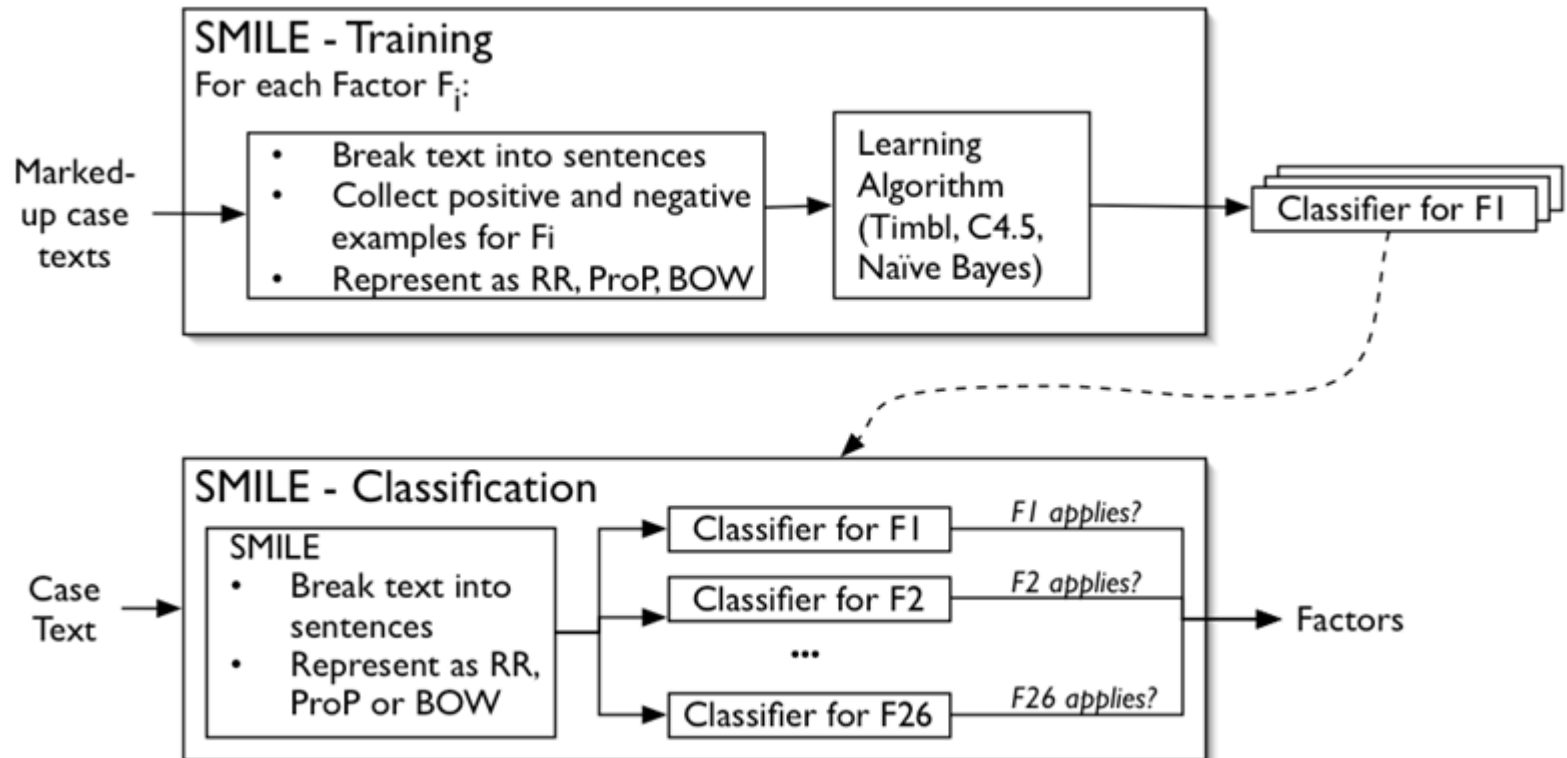
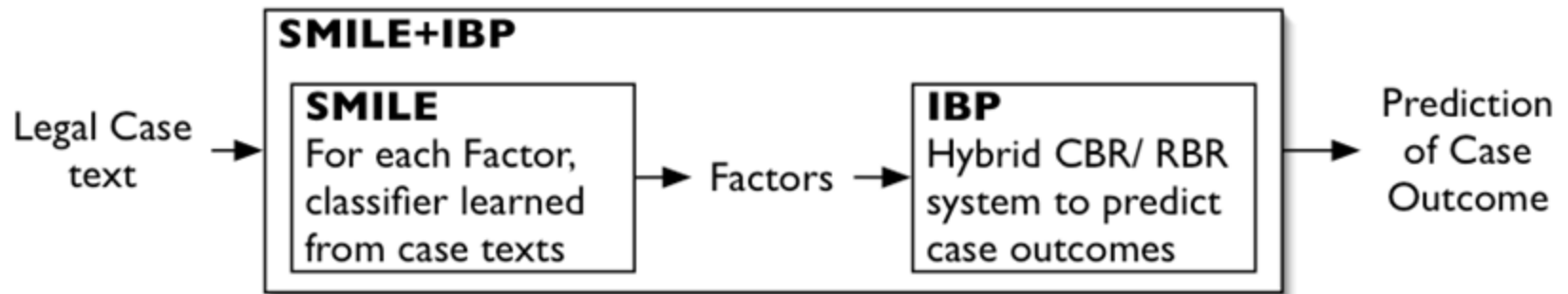
- KO Factors  $\equiv$  factors representing
  - behavior paradigmatically proscribed or encouraged under trade secret law and
  - for which probability  $P$  that a side wins when the Factor applies is at least 80% greater than the baseline probability  $BP$  of the side's winning.
    - $P$  = no. of cases where the Factor applies and the side won divided by no. of cases where the Factor applies.
    - $BP$  = no. of cases where the side won divided by no. of cases in collection
- If IBP:
  - finds KO-Factor accounts for outcome of counterexample, it deems the exception to be distinguishable and not a reason for abandoning its hypothesis.
  - can distinguish all counterexamples, then hypothesis is treated as confirmed.
  - cannot distinguish all counterexamples, it abstains for that issue and set of factors.
- IBP's KO-Factors :
  - Pro-P: F8 Competitive-Advantage, F26 Deception
  - Pro-D: F17 Info-Independently-Generated, F19 No-Security-Measures, F20 Info-Known-to-Competitors, F27 Disclosure-In-Public-Forum.
- IBP explains away:
  - *MBL* using KO-Factor F20 (d)
  - *CMI* using KO-Factors, F17 (d), F20 (d), and F27 (d)
  - Thus, the pro-plaintiff prediction is confirmed.

# Evaluation of IBP Algorithm

- 148 cases in CATO database, plus 38 new cases
- Experiments run in leave-one-out cross-validation; Relevance tested with McNemar's test
- Compare IBP with:
  - Baseline: predict majority class
  - Standard machine learning algorithms
  - Prediction based on CATO/Hypo relevance criteria



# Connecting predictive hypotheses to text: SMILE+IBP



# 3 text representations

---

Training instance for F4 Nondisclosure-Agreement:

“Diekman signed a nondisclosure agreement”

- Bag of words (BOW):
  - “a agreement Diekman nondisclosure signed”
- Roles Replaced (RR):
  - case-specific party and product names replaced with roles in case
  - “a agreement defendant nondisclosure signed”
- Propositional Patterns (ProPs):
  - combinations of words in 4 syntactic relationships: subject – verb, verb – object, verb – prepositional phrase, and verb – adjective
  - (defendant sign)(sign nondisclosure\_agreement)

# Research Questions

---

- I. Which representation enabled:
  - A. SMILE to do the best job of assigning Factors to cases as compared to the human assignments?
  - B. *SMILE+IBP* to do the best job of predicting outcomes of cases as compared to the actual outcomes?

*Hypothesis I.A:* Abstracting from names and individual entities in a case text to their roles in the case allows a learning algorithm to better generalize from training examples (i.e., RR > BOW).

*Hypothesis I.B:* Using some linguistic analysis to capture patterns of actions and negation preserves crucial information from the text and thereby leads to better classification (i.e., ProP > RR).

- II. How well did *SMILE+IBP* predict outcomes as compared to a baseline of informed guessing?

# Measures

---

- Assigning factors to test cases:
  - Precision  $\equiv$  proportion of correctly classified documents among all documents classified =  $(\text{correct}) / (\text{correct} + \text{false})$
  - Recall  $\equiv$  proportion of correctly classified documents out of all relevant documents =  $(\text{correct}) / (\text{correct} + \text{missed})$
  - F-Measure =  $(2 * \text{recall} * \text{precision}) / (\text{recall} + \text{precision})$
- Predicting outcomes of test cases:
  - Accuracy  $\equiv$  proportion of predictions that were correct =  $\text{correct} / (\text{correct} + \text{mistake})$
  - Coverage  $\equiv$  proportion of instances for which prediction was made =  $(\text{correct} + \text{mistake}) / (\text{correct} + \text{mistake} + \text{abstain})$
  - F-Measure(Pred) =  $(2 * \text{accuracy} * \text{coverage}) / (\text{accuracy} + \text{coverage})$

## SMILE+IBP Results: Comparing text representations

Text Representation	ProP	RR	BOW
Measure			
Avg. F-Measure	0.260	0.280	0.211
F-Measure(Pred)	0.703	0.600	0.585

Comparison of Text Representations	ProP v. BOW	RR v. BOW	Prop. v. RR
Measure			
Avg. F-Measure	(+) Sig.	(+) Sig.	(-) NOT sig.
F-Measure(Pred)	(+) Sig.	(+) NOT sig.	(+) Sig.

# SMILE+IBP: Comparing text representations; Discussion

---

- *Hypothesis I.A: classification*
  - RR > BOW? Confirmed
  - ProP > BOW? Confirmed
  - ProP > RR? No, but difference not significant
  
- *Hypothesis I.B: prediction*
  - RR > BOW? Yes, but difference not significant
  - ProP > BOW? Confirmed
  - ProP > RR? Confirmed

# Results: Comparing SMILE+IBP v. Informed Guessing

Baseline: biased coin flip

- In random experiment, predict that plaintiff wins with probability  $p = \text{no.-cases-won-by-plaintiff} / \text{total no.-cases} = 89/146 = 61\%$ .
- This strategy is preferable to always predicting majority class wins which ignores prior probability of defendant's winning.

Experiment II	SMILE+IBP	Baseline (biased coin flip)
Avg. F-Measure(Pred)	0.70	0.66

# Toward modeling normative hypotheses

---

## → 1. Propose rule for deciding current fact situation (cfs):

- Construct a rule 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 test case to probe if rule is too *broad*:

- Find/construct a test case that:
- emphasizes some normatively relevant aspect of the cfs and
- to which the proposed rule applies and assigns the same result as to the cfs, but
- where, given legal principles/policies, the result in test case is normatively wrong.

## → 3. Respond to test case:

### (3.a) Justify the proposed rule:

- Analogize the test case and the cfs and
- argue that they both should have the result assigned by the proposed rule. *Or*

### (3.b) Modify the proposed rule:

- Distinguish the test case from the cfs, argue that they should have different results and that the proposed rule yields the right result in the cfs, and
- add a condition or limit a concept definition so that the narrowed rule still applies to the cfs but does not apply to, or leads to a different result for, the test case. *Or*

### (3.c) Abandon the proposed rule and return to 1.

# Connecting normative hypotheses to legal texts: LARGO

## 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

IF

THEN

-Test

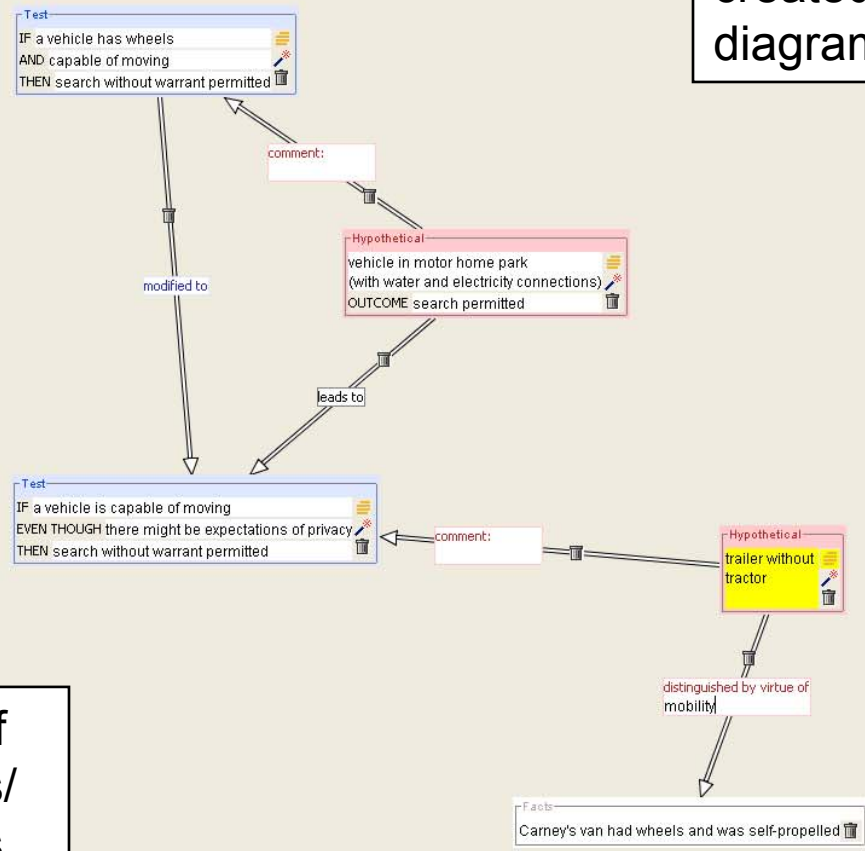
-Facts

Relations

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

## Palette of Elements/ Relations

## Student-created diagram



# Proposed approach to model semantic hypotheses

- Develop “hypothesis ontology” for knowledge transmission in civil litigation
  - explicit specification of concepts and relations in hypotheses
  - includes key concepts, such as “*communicated-with*” and “*during-interval*”
  
- Support users’ framing hypotheses in semi-structured “language”:
  - explicit specification of concepts and relations in hypotheses
    - there are documents of a particular kind,
    - satisfying particular time constraints,
    - satisfying particular social interaction constraints,
    - that refer to particular concepts or phrases of interest.
  - E.g., “there exist documents in which Alice *communicated-with* Bob *during-interval* 1976 to 1978 that *contain-keywords* {tobacco, children, advertising}”
  
- Induce (dynamic) social network based on document header info...and attorney feedback.
  
- Evaluate user-constructed hypotheses in terms of social network:
  - query leads system to
    - search for documents containing specified keywords,
    - expand results with plausible communications between the parties, and
    - refine results based on the specified interval and other factors.
  - cluster documents according to how they bear on hypothesis.

# Conclusions

---

- Ways to connect hypothesis-testing to legal texts.
  - Predictive, normative, or semantic hypotheses
  - Not just(!) retrieval but
    - retrieval + **problem solving**
    - reasoning with examples and counterexamples
- SMILE+IBP:
  - first & only program to reason automatically about legal case texts.
- Developing approaches to connecting normative or semantic hypothesis-testing to texts.