Doing Things with Factors

Vincent Aleven and Kevin D. Ashley
Intelligent Systems Program,
Learning Research and Development Center, and
School of Law
University of Pittsburgh
Pittsburgh, PA 15260
aleven+@pitt.edu, ashley+@pitt.edu

Abstract

We conducted an experiment to investigate whether a human tutor could employ the CATO model and instructional program to teach legal research and argumentation skills to beginning law students. The CATO model covers arguments comparing and contrasting cases in terms of factors, abstractions of facts that tend to strengthen or weaken a party's position on a legal claim. At the time of the experiment, the CATO program comprised tools and resources that help apply the CATO model to specific problems, most importantly, a case database and tools for retrieving, displaying, and comparing cases in terms of factors.

We compared human-led instruction with CATO against more traditional classroom instruction designed to teach the same skills, without the use of the CATO model or tools. The subjects were 17 first-semester students from the University of Pittsburgh Law School. We found that human-guided instruction with CATO was as good as classroom instruction. We also found that answers generated by the CATO program were scored higher than the students' answers, suggesting that the model can potentially be employed even more effectively to teach students. Examples drawn from protocols of CATO sessions illustrate that students can use the CATO model to guide and facilitate the construction of arguments and often go beyond the model's limitations, at least under the guidance of a human tutor.

1. Introduction

In this paper, we report the results of an experiment to investigate whether a human tutor could employ the CATO program and its model of case-based legal argumentation to instruct law students. CATO is an intelligent tutoring system designed to teach law students to make legal arguments for and against results in problem disputes by comparing and contrasting the problem to precedents. At the time of the experiment, it provided tools and resources, but did not yet do any active tutoring. Under the guidance of a human tutor, beginning law students carried out legal research and argumentation tasks, using the CATO model as an organizational framework for representing a problem's factual strengths and weaknesses, finding cases, evaluating cases' relevance, and making arguments with cases. Instead of the standard law library tools to track down cases, students used CATO's case retrieval and display tools.

CATO's model and instructional environment channels students into retrieving and comparing cases in terms of factors. Factors, abstractions of facts that tend to strengthen or weaken a party's position on a legal claim, were introduced in HYPO's Dimensions [Ashley, 1990; 1991] as an AI/CBR knowledge representation device for indexing cases and abstractly summarizing the legally relevant facts of cases. Factors may be thought of as fact-oriented reasons for or against deciding a claim in favor of a plaintiff. A factor may be neither a necessary nor a sufficient reason for deciding a claim; the presence of a factor, however, *ceteris paribus* makes a case stronger or weaker for plaintiff. Generally, a case or problem presents a conflict among factors, in that some factors favor the plaintiff, others favor the defendant. To make arguments about how to resolve the issue, it is useful to compare and contrast the problem to cases that presented similar sets of factors.

Although factors have appealed to a wide range of researchers as a useful knowledge representation device (see, for example [Clark, 1990; Sanders, 1991; Hage, 1993; Loui, et al., 1993; Murbach and Nonn, 1993; Rissland, Skalak and Friedman, 1993; 1994]), the AI and Law research community lacks much experience in how human subjects interact with factors. Like most AI knowledge representation tools, employing factors is something of a double-edged sword. On the one hand, these stereotypical abstractions enable a program to compute the relevance of cases according to a variety of apparently useful definitions as in [Ashley, 1990; Ashley and Aleven, 1994] and to generate various types of case-based arguments. Case representations based on factors have also been used to retrieve, semi-automatically, small collections of cases that can be used to illustrate argumentation issues [Ashley and Aleven, 1992]. We had observed (but needed to confirm experimentally) that CATO's use of factors of-

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

© 1995 ACM 0-89791-758-8/95/0005/0031 $1.50

This work is supported by an NSF Presidential Young Investigator Award and grants from the National Center for Automated Information Research, West Publishing Company and Digital Equipment Corporation.

fered certain instructional benefits, for example, help students develop some interpretations of what a case means, focus students on the importance of comparing and contrasting cases in terms of factual strengths and weaknesses, and help students make initial judgments of the relevance of cases [Aleven and Ashley, 1994].

On the other hand, as stereotypes, factors and the relevance criteria based on them are subject to limitations in terms of what they leave out, for instance, knowledge related to what the factors mean and why they matter. Currently, the factors' applicability conditions are not represented in a form that CATO can understand. Also, there is no representation of a factor's theoretical legal significance, for example, explanations linking factors to more abstract legal knowledge such as policies and purposes underlying a legal domain. Berman and Hafner have predicted that "case-based reasoners that incorporate teleological arguments will prove more useful to less skilled advocates or legal educators desirous of enhancing the advocacy skills of their students" [Berman and Hafner, 1993]. Thorne McCarty has also criticized a focus on factors as relying too much on "fixed features" and ignoring the importance of constructing theories to justify positions, theories that relate underlying concerns of the law with precedents and a proposed result. [McCarty, 1994]. As a tutoring device, there also is a fear that students will be tempted to rely on factors too much, to apply them mechanically or uncritically as symbols, forgetting what the factors mean and ignoring their limitations. As Berman has said, "CBR may prove useful as electronic trainers to enable fledgling lawyers to improve their basic skills at using cases. But like ball machines that cannot replicate a talented player's probing of an opponent's weaknesses, these legal trainers will not prepare lawyers to respond intelligently to the skilled lawyer's exploitation of the indeterminacy that inheres within legal precedents." [Berman, 1991].

Given the interest in factors, both positive and sceptical, and our own desire to insure that a human tutor could teach valuable lessons with CATO's factor-oriented model, it was important to document empirically how human subjects reacted to CATO's use of factors. In this paper, we describe the CATO program as students saw it at the time of the experiment, describe the experiment, report and discuss its findings, and, finally, present examples drawn from the experimental protocols which illustrate ways in which students used factors.

2. The CATO model and instructional program

In this section, we describe the CATO program by illustrating how one law student used its tools and resources to analyze a problem. The goal in the CATO project is to develop an instructional environment that students can use to practice legal research and argumentation tasks and develop certain of the skills involved. In particular, the instructional goal is for students to learn to analyze a problem, select relevant cases based on a comparison of the cases' factual strengths and weaknesses, and develop arguments about the problem supported by the most relevant cases. In addition, the goal is for students to learn to frame effective queries for automated case law databases, in particular, Westlaw's natural language query system, WIN [Croft and Turtle, 1992].

The CATO model of case-based argumentation provides a general plan describing an argument's overall structure and a set of argument moves to employ as building blocks of an actual argument [Aleven and Ashley, 1993; 1994]. The model covers argument moves such as analogizing the problem to cases with a favorable outcome, citing cases to emphasize strengths or to downplay weaknesses, and covering the opponent's bases (i.e., the weaknesses present in the problem). It also includes responses to such arguments by distinguishing and citing counterexamples. The CATO model provides a set of relevance criteria that specify which cases are the best to use in each argument move. Core elements of the model, most importantly, the use of factors to represent cases and define relevance criteria, originate in HYPO's model of reasoning with cases [Ashley, 1990; 1991].

The CATO instructional environment is shown in Figure 1. Currently, it provides tools and resources that help students to apply the model to a particular problem, most importantly, a case library and tools for retrieving, displaying, and comparing cases in terms of factors. CATO's case library contains, for each of 45 trade secrets cases, a (pre-stored) list of applicable factors and a squib (i.e., a narrative summary). The CATO query language enables students to retrieve cases with any boolean combination of factors. CATO can display the factors or squib of a case, and show a comparison of the factors of a retrieved case and those of the problem. Currently, CATO does not communicate to students the argument plan, argument moves, and relevance criteria, but it will in the future.

We presented "Steve", a third year law student with no prior knowledge of the CATO system, with a problem based on *Mason v. Jack Daniel Distillery,*[1] an Alabama

[1] In 1980, a restaurant owner named Mason developed a combination of Jack Daniel's whiskey, Triple Sec, sweet and sour mix, and 7-Up to ease a sore throat. He promoted the drink, dubbed "Lynchburg Lemonade" for his restaurant, "Tony Mason's, Huntsville", served it in Mason jars and sold T-shirts. Mason told the recipe only to his bartenders and instructed them not to reveal the recipe to others. The drink was only mixed out of the customers' view. The drink comprised about one third of the sales of alcoholic drinks. Despite its extreme popularity, no other establishments had duplicated the drink, but experts claimed it could easily be duplicated.

In 1982, Randle, a sales representative of the Distillery, visited Mason's restaurant and drank Lynchburg Lemonade. Mason disclosed part of the recipe to Randle in exchange, Mason claimed, for a promise that Mason and his band would be used in a sales promotion. Randle recalled having been under the impression that Mason's recipe was a "secret formula".

Randle informed his superiors of the recipe and the drink's popularity. A year later, the Distillery began using the recipe to promote the drink in a national sales campaign. Mason did not participate in the promotion or receive other compensation.

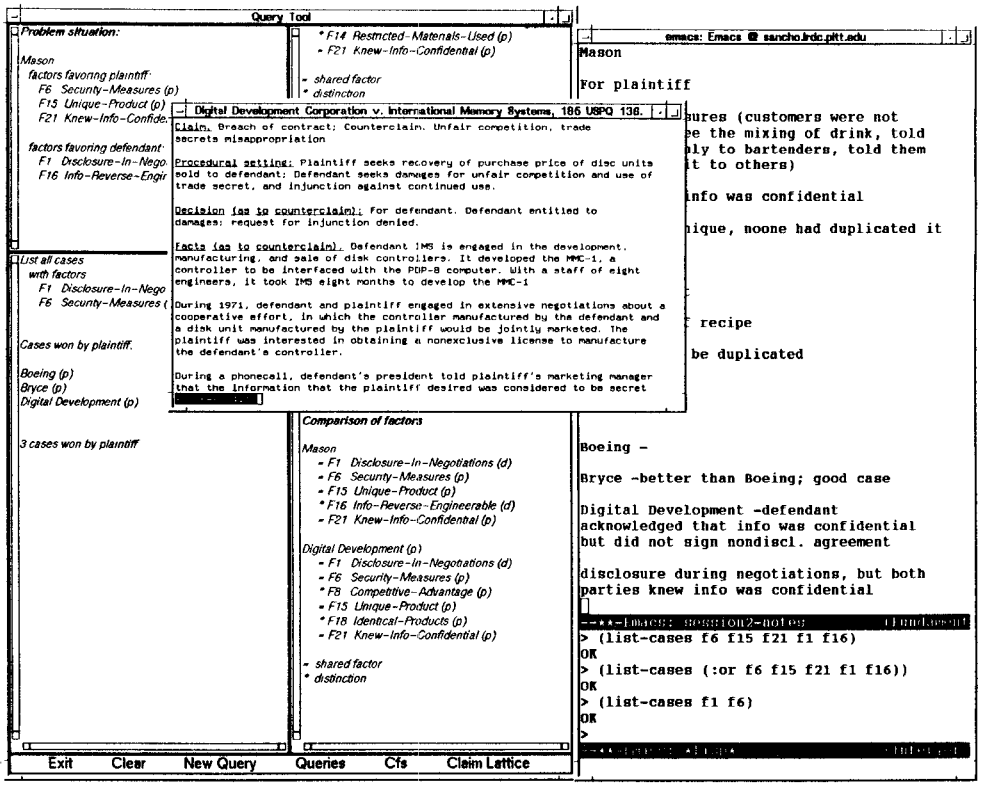

Figure 1: The CATO interface. The factors that the student identified for the problem situation ***(Mason)*** are displayed in the top left window. There are windows for entering queries (bottom right), displaying the result of the last query (bottom left), and displaying a case's factors or a comparison with the factors of the problem (middle window). Queries are expressed in a Lisp-like syntax. CATO displays squibs (short summaries of cases) in separate pop-up windows. Students can also request to see a history of their queries (middle window) and can make notes about the cases they have seen in the top right window.

case which involved a claim for trade secret misappropriation (see [Aleven and Ashley, 1993] for a more complete analysis of the *Mason* case). We introduced Steve to some of the trade secrets factors and cases in the CATO database, and showed him how to use the CATO tools. Finally, we asked him to make an argument on behalf of the plaintiff in *Mason*, supported by cases he would have to find and retrieve from CATO's database. Figure 2 summarizes the strategy he adopted on his own, without prompting from us. First, he identified the strengths and weaknesses of the plaintiff's side in terms of factors. (We describe that process in Section 4.1.) Then he formulated an argument for the plaintiff by identifying three issues and searching CATO's database for cases to resolve them. He was able to formulate each issue in terms of factors and to fashion queries to CATO specifying factors and outcomes that retrieved cases should, may, or should not have. He also used factors to make initial judgments of the retrieved cases' relevance to the issue. We illustrate examples of his reasoning and queries for two of those issues in Sections 4.3 and 4.4. After carrying out this strategy, Steve was convinced that "the plaintiff has now made a solid argument on their case."

3. An Experiment with CATO

In the late fall of 1993, we conducted an experiment with 17 first-semester law students. The goal of the experiment was to get an early indication of the instructional effectiveness of the CATO environment, in teaching argument-making and case-finding skills to beginning law students. A second goal was to gather more detailed information about how students interact with the CATO model and tools. The experiment was designed to compare instruction with CATO against more traditional methods of teaching the same material. Specifically, the plan was to compare students' abilities after roughly equal time periods of more traditional legal methods classroom instruction and instruction using CATO. This was meant to be a formative evaluation [Shute and Regian, 1993], designed

Using Cato to analyze the ***Mason*** problem and construct an argument for the plaintiff, student:

1. Identified factors present in ***Mason*** and whether they were strengths or weaknesses of plaintiff's side. (For a description of this process see Section 4.1.)
2. Formulated an argument for the plaintiff by identifying issues and searching for cases.

   2.a **Issue 1:** How to blunt the defendant's strengths using (some of) plaintiff's strengths? Defendant's strengths were that the plaintiff disclosed its secrets to the defendant in negotiations (Factor F1) and the product was easy to reverse engineer (F16).

   Queries:

   (1) List all cases with the same factors as ***Mason*** (returned no cases);

   (2) Relaxed query 1: List all cases won by plaintiff with one or more of the pro-plaintiff factors and all of the pro-defendant factors present in ***Mason*** (returned no cases);

   (3) Relaxed query 2: List all cases won by plaintiff with one or more of the pro-plaintiff factors and one or more of the pro-defendant factors present in ***Mason*** (5 cases);

   (4) Restricted query 3, by adding the constraint that the retrieved cases have no other factors (1 case).

   Lead to: ***Digital Development*** case but there was a distinction: In ***Mason***, plaintiff's secret could be reverse engineered with some effort (i.e., pro-defendant factor F16: Info-Reverse-Engineerable applied to ***Mason*** but not to ***Digital Development***).

   2.b **Issue 2:** How to blunt the effect of that distinction, factor F16: Info-Reverse-Engineerable?

   Queries: List all cases with factor F16 (returned 4 cases, 1 won by plaintiff).

   Lead to: ***American Precision*** case where plaintiff won despite F16. (For student's argument with ***American Precision,*** see Section 4.4.)

   2.c **Issue 3:** How to blunt defendant's remaining strength, factor F1? In ***Mason*** plaintiff made disclosures to defendant (factor F1) without securing a nondisclosure agreement (factor F4 Agreed-Not-To-Disclose, absent in ***Mason***).

   Queries:

   (1) List all cases with factor F1 (returned 8 cases, 4 won by plaintiff).

   (2) Restricted query 1: List all cases with factor F1 but without factor F4 (6 cases, 2 won by plaintiff).

   (For a discussion of student's reasoning and queries, see Section 4.3.)

   Lead to: ***Digital Development*** and ***Space Aero*** cases where plaintiff won despite disclosures to defendant without nondisclosure agreement.

3. Composed argument for plaintiff from notes summarizing above issues and results of queries.

Figure 2: Observed student's strategy using CATO to analyze ***Mason*** problem and construct an argument for the plaintiff.

to gather information that could aid in the further development of the CATO program.

The subjects were recruited by advertizing in the student newsletter of the University of Pittsburgh Law School. They were randomly assigned to a control group of 10 students and an experimental group of 7 students. All subjects were given an introductory lecture about trade secret law, a subject not taught in the first year legal curriculum. The instructor discussed parts of the Restatement of Torts (2nd), section 757, which includes a list of six factors for trade secrets law. All of the subjects also participated in a hands-on session led by a Westlaw instructor explaining how to use Westlaw and WIN.

All subjects then took a pre-test take-home examination comprising two types of questions: (1) argument-making questions designed to assess their abilities in making legal arguments citing cases and (2) case-finding questions to assess their skills in drafting queries to retrieve relevant cases with WIN. The purpose of the pre-test was to verify that the students in the control group and experimental

group were, on the average, of equal ability —as one would expect, given that the students were randomly assigned to each group.

The control group received two two-hour sessions conducted by two legal methods instructors in the Law School's computer laboratory room. Each session began with a classroom lecture, one on how to analyze a legal fact situation and conduct legal research about it using Westlaw and WIN. The other lecture dealt with how to make a legal argument. The list of trade secrets law factors presented in the Restatement of Torts (2nd) played a rather central role in one of the lectures. Following each lecture, the students were assigned specific trade secrets problems and the task of finding relevant cases using WIN and outlining arguments about the problem. In the second session, the students made brief oral arguments for a side based on the cases they had found and responded to those arguments on behalf of the opponent. Although the control group sessions were not part of the regular methods curriculum, but were organized especially for the experiment, we believe that the control group instruction constitutes a reasonable attempt to teach, in a traditional way, the same skills as are targeted by the CATO instruction.

Each student in the experimental group received two two-hour sessions of instruction using CATO. During these sessions, he or she learned to use the CATO tools and analyzed two trade secret problem situations, including the *Mason* problem, guided by a human tutor, Kevin Ashley. Each session was attended by a pair of students. The students received practice in identifying a side's strengths or weaknesses in terms of factors, citing representative cases, resolving conflicts among factors, distinguishing unfavorable cases, citing counterexamples to trump an opponent's cases, and covering the opponent's bases to overcome his strengths. The students dictated their queries to Vincent Aleven who translated them into CATO's query language. (In fact, he handled all interaction with the program.) We made video or audio recordings of each of the CATO sessions, 11 in total, 9 sessions as part of the experiment, and 2 additional sessions prior to the experiment.

Following the instructional sessions, the students took a take-home post-test containing the same kind of argument-making and case-finding questions as the pre-test. The purpose of the post-test was to assess the students' skills after their respective instruction, so that the relative effectiveness of the two types of instruction could be compared. In particular, the question was whether there was a significant difference in post-test scores, between experimental group and control group, that was not seen on the pre-test (or vice versa).

The students' responses on the pre- and post-tests were graded in blind tests by two legal methods instructors (not those who had conducted the control group instruction) and three West Publishing Company reference attorneys. The West graders graded only the students' WIN queries. None of the graders was familiar with the CATO model. For the argument-making questions, we included in the materials to be graded a set of answers generated by CATO. We did not tell the graders that this output was computer-generated. We provided a set of grading criteria which were compatible with the CATO model but were not phrased specifically in terms of the CATO model. (Since the control group instruction was not based on the CATO model, it would not have been a fair test to use the model as the standard to evaluate student performance.) The grading criteria rewarded students' identifying important issues and factual strengths and weaknesses in a problem, appropriate selection of cases to cite in an argument (subdivided into further criteria), students' correctly drawing analogies, distinguishing cases, and using counterexamples. The grading criteria also included more general criteria such as scope of analysis, context sensitivity, and effectiveness. A slight penalty was assessed for "negatives" such as "so what?" differences, conclusory statements, repetition, etc. The students' written WIN queries were graded based on two criteria: Whether the query captured important issues raised by the problem, and whether it was (or would be) effective in retrieving relevant cases.

We conducted a policy capturing study, asking the West graders to explain why they felt certain queries were better than others. This resulted for each grader in a list of criteria, which we believe will be useful to help students in learning to use WIN effectively.

The quantitative results of the experiment are shown in Table 1 and may be summarized as follows: (1) On both the pre-test and the post-test, the control group's scores were slightly higher than those of the experimental group. The differences, however, were not statistically significant. Note that the pre- and post-test scores for the argument-making questions are not directly comparable because they were scored by different graders and the tests comprised different mixes of questions. Thus, the fact that the post-test scores for both groups on the argument-making questions were somewhat lower than the pre-test scores does not convey any useful information. (2) The CATO-generated answers to the argument-making questions were scored significantly higher than the students' answers on both the pre- and post-tests. (3) There is a slight increase in the student scores for the case-finding questions (involving WIN) on the post-test as compared to the pre-test. The difference is statistically significant. Since the same graders graded the pre- and post-test case-finding questions, these scores are comparable.

We interpret the first result as confirming that the CATO instruction (as guided by a human tutor) is as effective as the more traditional classroom instruction. Although the results do not indicate that the CATO instruction was more effective than the more traditional instruction, the second finding suggests that the CATO model has the potential to be more effectively employed than it was in this experiment. Apparently, the experimental group students did not learn everything there was to learn from CATO's argumentation model. The model enables CATO to generate significantly better answers, but the experimental group did not learn the skills necessary to generate answers as good as CATO's.

	Pre-test			Post-test		
	argument-making	argument-making and case-finding	WIN queries	argument-making	argument-making and case-finding	WIN queries
Experimental Group Average	52	45	52	41	42	57
Control Group Average	56	50	51	46	45	57
CATO score	68*			78*		

Table 1: Percentage of maximum score on pre-test and post-test. * denotes highest score.

The third finding gives some evidence that both groups did, in fact, learn something about using WIN effectively by virtue of their respective instruction. Since the pre- and post-test scores for the argument-making questions are not directly comparable, we cannot similarly confirm that both groups learned something about argument-making from their respective instruction. As noted above, this was because different graders graded the pre- and post-tests, and because the tests may not have been equally difficult.

Given that students used CATO only under the continuous guidance of a human tutor, it is clear that this experiment does not support any conclusions about the effectiveness of CATO as a stand-alone tutoring system. We plan to undertake such an experiment after further developing the CATO program.

4. Examples of students' using factors in legal research and argumentation

In this section, we show examples, taken from the protocols of the CATO sessions, that illustrate ways in which students used the CATO model and tools to guide and facilitate their legal research activities. As discussed in the introduction, CATO's case representations based on factors leave out certain knowledge related to what the factors mean and why they matter. This includes abstract legal knowledge about the policies and purposes underlying the given area of the law, knowledge about the corporate world, and common sense knowledge. However, it is important that students draw on this knowledge when making arguments with cases and do not treat factors as icons or manipulate them mechanically.

Students can be expected to have at least some of this background knowledge; we took various measures to make more of it available. All students received a lecture about trade secrets law. Some of the background knowledge was available to students in the squibs and even in the names of factors. We impressed upon students that it is important to know the specific case facts behind the factors. We introduced the concept of factors by illustrating how factors correspond to the factual strengths and weaknesses that students identified for a few sample cases. Also, we stressed the fact that not every instance of a factor is the same: Even if cases share a factor, they may differ in how extreme their facts are with respect to that factor[2]. Finally, we repeatedly encouraged students to read the squibs of the cases.

The question is whether these measures ensure that students draw on their background knowledge appropriately while making arguments with cases, even if the CATO model does not provide explicit guidance or examples in this respect. In presenting the examples, we pay particular attention to whether students were able to integrate into their arguments knowledge about the meaning of factors and the other types of background knowledge.

4.1 Students used factors to represent the strengths and weaknesses that a problem presents.

After reading a short description of the facts of the problem situation, students identified factual strengths and weaknesses, that is, facts that would tend to help plaintiff's or defendant's case. The tutor then helped them select the corresponding factors. For example, we asked Steve to analyze the *Mason* problem and identify any applicable factors (see Section 2). In CATO's database, *Mason* is indexed by the following factors: the plaintiff disclosed its secrets to the defendant in negotiations (Factor F1) and the product was easy to reverse engineer (F16), two factors strengthening the defendant's position. On the other hand, in plaintiff's favor, it took some security measures (F6), its product was unique (F15), and the defendant's representative knew the information was confidential (F21).

A student may agree with CATO's set of applicable factors, but he need not. As it happened, Steve agreed with the above set of factors for *Mason*, but he also found that two additional factors applied and identified two sets of facts he regarded as relevant but for which CATO offered no factors. In situations where a student does not agree

[2] In HYPO, but not in CATO, the magnitude of each factor was represented explicitly [Ashley, 1990; 1991].

with the set of factors associated with the case in CATO's database, but decides that some other factors in CATO's database apply, we simply use the student's list as the canonical analysis of the problem. If the student decides that factors apply which are not known to CATO, we can only ask him to list his factors for future reference.

4.2 Students used factors to implement a general research or argumentation strategy.

Students used the CATO query language to implement general research strategies for finding cases to cite in an argument. Even though the strategies were not part of the CATO model and we did not communicate them to the students, it turned out that many searches expressed in terms of factors bore a close connection to CATO's argumentation strategies.

We observed four general research strategies. First, students often looked for cases that have similar sets of factors as the problem. Students executed queries for cases with the same set of factors as the problem, or, when this failed, for cases that have at least one factor in common with the problem. Second, some students focused on the favorable factors that are present in the problem, evidently with the intention to argue for a favorable result on the basis of their strengths. Others students focused on finding cases that share (some of) the unfavorable factors with the problem, intent on blunting the the opponent's strengths. Finally, some students searched for cases that combine some of the unfavorable factors and some of the favorable factors. These queries turned out to be very effective. Students also specified that retrieved cases should not have certain distinctions, or no distinctions at all. We believe that practice with these general research strategies will turn out to be valuable for students, even though the query languages of commercially available retrieval systems such as WIN or Westlaw differ substantially from CATO's.

A sequence of queries in Steve's example illustrates how these strategies can be used, with some trial and error, to find relevant cases (see Figure 2, issue 1). In order to retrieve cases to cite for the plaintiff in the *Mason* problem, he first tried to find cases with the same set of factors as *Mason*. When this failed, he tried to find cases with combinations of the favorable and unfavorable factors. His main concern was to find the cases needed to support an argument that the plaintiff could win in spite of the unfavorable factors. Having found some cases, he added a constraint (in the fourth query) that the cases should not have any factors that are not present in the problem, so as to filter out any distinguishable cases.

4.3 Students used factors to formulate issues and focus research in a given problem situation.

Steve's example also illustrates how students used factors to define issues and focus the search for cases. While analyzing the *Mason* problem, he defined three issues in terms of factors (see Figure 2). He further refined the third issue in a very interesting way, applying knowledge about the meaning of factors that is not represented in the CATO model. Steve needed to find a way to argue that plaintiff in *Mason* could win in spite of a weakness in its position: Mason had disclosed his alleged secret to the defendant during negotiations (pro-defendant factor F1 Disclosure-In-Negotiations). Steve first ran a query for cases that have this factor and found several cases in which the plaintiff won despite the presence of this factor:

```
> (list-cases f1)

List all cases
   with factor
      F1 Disclosure-In-Negotiations (d).

Cases won by plaintiff:
Boeing (p)
Bryce (p)
Digital Development (p)
Space Aero (p)

(4 cases won by defendant omitted.)
```

Steve realized, however, that the cases won by plaintiff may be distinguishable in a potentially troublesome way. He knew that trade secrets law allows plaintiffs to communicate information to others pledged to secrecy. The plaintiff in *Mason*, however, had failed to take the precaution of securing a nondisclosure agreement from the defendant (factor F4 Agreed-Not-To-Disclose, not present in *Mason*). Therefore, it was important to know whether plaintiff could win despite such lack of care. Steve therefore refined his query to select only the cases in which F4 did not apply:

```
> (list-cases f1 :not-any f4)

List all cases
   with factor
      F1 Disclosure-In-Negotiations (d)
   but without factor
      F4 Agreed-Not-To-Disclose (p).

Cases won by plaintiff:
Digital Development (p)
Space Aero (p)

(4 cases won by defendant omitted.)
```

The query returned two cases in which the plaintiff won, confirming Steve's hypothesis and giving him the cases he needed to support his argument. It is interesting to note that in refining his initial query, Steve applied knowledge about trade secrets law that is not represented explicitly in the CATO model. Nonetheless, CATO's vocabulary of factors may have helped him frame the issue; certainly, CATO's database and query language based on factors enabled him quickly to find some relevant cases.

4.4 Students used factors to make initial assessments of what cases mean and whether they are relevant.

Students often used the factor representations of cases to make initial judgments of what the cases mean and how they could be used in an argument. For this purpose, they often used CATO's tool for comparing the factors of a re-

trieved case to those of the problem. CATO marks the shared factors and the distinctions, as in the following comparison of the *Eaton* case and the *Motorola* problem:

```
Motorola
    * F2 Bribe-Employee (p)
    = F4 Agreed-Not-To-Disclose (p)
    = F5 Agreement-Not-Specific (d)
    = F6 Security-Measures (p)
      F10 Secrets-Disclosed-Outsiders (d)
    = F17 Info-Independently-Generated (d)
    = F20 Info-Known-To-Competitors (d)

Eaton (d)
    = F4 Agreed-Not-To-Disclose (p)
    = F5 Agreement-Not-Specific (d)
    = F6 Security-Measures (p)
    * F16 Info-Reverse-Engineerable (d)
    = F17 Info-Independently-Generated (d)
    = F20 Info-Known-To-Competitors (d)

= shared factor
* distinction
```

Distinctions are those unshared factors that push toward an opposite result in the two cases. The comparison of factors gives an impression of how similar the case is to the problem and how the case could be used in an argument. In the CATO model, an arguer analogizes the problem to a case by citing the shared factors, and distinguishes it by pointing out the distinguishing factors (for information about other argument moves, see [Aleven and Ashley, 1993]). If a case looks promising on the basis of its factors (as does *Eaton* in the *Motorola* problem), it is important to read it and know the specific facts behind the factors. As one student said: "[A]s detailed as these factors are, you still need to read very carefully, because even a factor [such as] Unique-Product, there's just such a wide array of what that word and that factor means." If a case does not look promising on the basis of its factors, however, there is little point in reading it. This way, queries and initial relevance judgments based on factors enable students judiciously to select cases for reading, and to avoid wasting time reading cases that end up not making a contribution to the argument.

On several occasions, students were able to improve and refine relevance judgments based on factors, by taking advantage of knowledge not represented as factors. Often, they tried to "account for the distinguishing factors," as one student put it. In the *Eaton/Motorola* example, the students argued convincingly that an apparent distinction was at most a minor difference, by reasoning about the meaning and impact of a distinguishing factor in light of other factors present in the problem. (We found similar examples in at least three other transcripts.) Looking for cases to cite on behalf of the defendant in *Motorola,* the students found *Eaton.* The comparison of factors shown by CATO (see above) indicates that although *Eaton*'s set of factors is quite similar to *Motorola*'s, plaintiff can distinguish it: In *Eaton,* but not in *Motorola,* plaintiff's information could be (legally) discovered through reverse engineering (i.e., by examining or analyzing plaintiff's product), factor F16. However, the students made a reasonable argument that this factor was not a major distinction. They stated that "*Eaton* would be a fairly strong case for the defendant because you've only got one distinguishing [factor] and the fact that it's reverse engineerable [F16] is actually not a big problem because the information was independently generated [F17], the information was known to [competitors (F20)] ... it's just not a big distinction." In other words, the students reasoned that when certain information is generally known in an industry (F20), the fact that it can also be discovered by reverse engineering (F16), hardly makes it any more widely available than it is already.

This example also illustrates a second way in which students were able to eliminate seeming distinctions: applying their knowledge about the corporate world, the students argued that the specific facts behind a distinguishing factor simply constitute normal practice and should not be interpreted as a strength (or factor) for the plaintiff. The comparison of the *Motorola* and *Eaton* factors indicates that plaintiff may distinguish *Eaton*: In *Eaton,* the defendant offered former employees of the plaintiff substantial bonuses or salary increases in order to induce them to switch employment (factor F2 Bribe-Employee), raising the suspicion that they were offering employees a bribe in order to bring trade secrets. However, the students did not think that this interpretation of the *Motorola* facts was warranted. Earlier during the session, when comparing *Motorola* to a different case, they had stated: "I think it would be fairly easy for the defendant to say ... that in a market-oriented situation bonuses and salaries were being made to encourage ... the defendants to come over to the company. Not necessarily for any devious means but just because they were skilled workers ..."

The following example shows that a third-year student (Steve) was able to improve upon arguments based on the CATO model, by analogizing cases at a more abstract level than factors and by relating the factors of a case to the court's rationale for deciding it. Arguing on behalf of the plaintiff in *Mason,* Steve looked for cases that he could use to blunt one of defendant's strengths: The fact that plaintiff's allegedly secret information could be discovered by reverse engineering plaintiff's product (pro-defendant factor F16 Info-Reverse-Engineerable) (see Figure 2, issue 2). He found *American Precision,* a case in which the plaintiff won despite the presence of this factor:

```
Mason
    * F1 Disclosure-In-Negotiations
      F6 Security-Measures (p)
      F8 Competitive-Advantage (p)
      F15 Unique-Product (p)
    - F16 Info-Reverse-Engineerable (d)
    - F21 Knew-Info-Confidential (p)

American Precision (p)
    * F7 Brought-Tools (p)
    - F16 Info-Reverse-Engineerable (d)
    - F21 Knew-Info-Confidential (p)

- shared factor  * distinction
```

The plaintiff in *Mason* can cite *American Precision* to support an argument that it is not fatal to plaintiff's trade secrets claim that the information could be discovered by reverse engineering. However, the defendant can respond by distinguishing *American Precision*, for example by pointing out that in *American Precision*, but not in *Mason*, the defendants took documents containing plaintiff's secret information (factor F7).

Steve was able to made a stronger argument, namely, by applying the court's rationale in *American Precision* (as summarized in the squib), to *Mason*. In *American Precision*, he stated, the defendants acted fraudulently in taking product development information (F7) while knowing that plaintiff considered the information to be confidential (F21). The court held that "It did not matter that [defendant's information] could have been developed in a lawful way. ... [W]here ... [information] has been unfairly acquired, it will be afforded protection as a trade secret." In *Mason*, Steve argued, there was evidence of similar fraudulent use of information known to be confidential (F21). Applying the rationale of *American Precision* to *Mason*, he argued that the plaintiff in *Mason* should win, regardless of the fact that its information may have been reverse-engineerable (F16). This is a stronger argument than saying that this factor is not necessarily fatal.

These three examples show that students were able to use factors to make initial judgments of cases' relevance, but were also able to go beyond the comparisons based on factors by applying knowledge about what the factors mean and why they matter.

4.5 Students used factors to formulate and test theories about trade secrets law.

The following two examples illustrate that students on several occasions made predictions about the relative importance of factors, in effect formulating a general theory about trade secrets law, expressed in terms of factors. They then tested their theories by retrieving cases from the CATO database. In both examples, students predictions seem to be prompted by reasoning about why certain factors matter.

One student formulated and tested a theory based on factor F17 Info-Independently-Generated. This factor applies when the defendant has developed a product to compete with plaintiff's through its own independent efforts, without using the information that plaintiff claimed to be its trade secret. The student predicted that where this factor applies, defendants would win, "because the whole idea of trade secrets is [to protect] against misappropriation and use [of confidential information] in developing a similar product." When the tutor challenged her to test her theory, the student executed a query for cases with factor F17 Info-Independently-Generated. The defendant won in all four cases that were retrieved, which confirms that the factor indeed is a strong factor for the defendant.

Another student predicted that certain factors would outweigh others; he was able to refine his theory when he found cases that were inconsistent with it (see Figure 3 — references to factors have been marked, using square brackets). Looking at the list of applicable factors that he had identified in the *Motorola* problem, he noticed an apparent contradiction, which led him to hypothesize that where plaintiffs had disclosed their secrets to outsiders (factor F10) or where the information was otherwise know to competitors (F20), plaintiffs would lose regardless of whether they had taken security measures (F6), because "once the information is not secret, what's the point of security measures?" When the tutor challenged him to test his theory, he retrieved—to his evident surprise—several cases that were inconsistent with his prediction. On the tutor's instigation, he inspected the factors of one of the retrieved cases and discovered that his theory could possibly be salvaged. He could have run a new query to test the revised theory, but as there was little time left, the tutor moved on to other things.

Formulating and testing theories based on a rationale that relates the theory to the purposes underlying the domain is a very important aspect of legal scholarship and is closely related to legal argumentation. The two examples illustrate that even with a fixed vocabulary of factors it is possible to provide some support for useful exercises in theory formulation and testing. The knowledge needed to reason about the relative importance of factors is not represented in the CATO model. But by providing a vocabulary of factors, CATO seems to inspire such reasoning on the part of students, and by providing a case database and query language based on factors, CATO makes it much easier to test the resulting theories.

Having gone through a few cycles of the theory-testing process in the CATO environment, students may be more able to recognize their own spontaneous generalizations about a legal domain as theories that can be tested against cases. Exercises like the ones illustrated here would be far more difficult to achieve with a full-text retrieval system. Such efforts would require far more time, or would even flounder with the complexity of figuring out what the retrieved cases mean. CATO reduces some of this complexity, because in the CATO database, cases are indexed by the same vocabulary of factors that students use to express theories (see also [Aleven and Ashley, 1994]).

Student ... when those two contradict, the secrets being disclosed [F10] would cancel out the ... they would make it irrelevant that there were security measures [F6], because the only point for security measures is to keep the information secret and once the information is not secret, what's the point of security measures?

Tutor How would you test that?

Student As far as, if that's true, you would do a search for f6 and f10 or f20. .... According to my theory, all those cases should go for the defendant.

```
> (list-cases f6 (:or f10 f20))

List all cases
   with factor
      F6 Security-Measures (p)
   and with one or more of factors
      F10 Secrets-Disclosed-Outsiders (d)
      F20 Info-Known-To-Competitors (d).
```

```
Cases won by plaintiff:
Boeing (p)
Data General (p)
Drill Parts (p)
FMC (p)

Case won by defendant:
Eaton (d)
```

Tutor Now *Eaton* went for the defendant ... How about those others?

Student That boggles my mind!

Tutor How would you resolve this?

Student I would read the cases and see what's going on.

Tutor Let's just take a look at *Data General*, that's a good one.

```
Data General (p)
  F6 Security-Measures (p)
  F10 Secrets-Disclosed-Outsiders (d)
  F12 Outsider-Disclosures-Restricted (p)
  F14 Restricted-Materials-Used (p)
  F18 Identical-Products (p)
```

Tutor ... Can you save your theory?

Student Well, actually, somewhat in that secrets disclosed to outsiders means that it might be a limited disclosure uh within your factor.

Tutor That is captured by the addition of another factor in this *Data General* situation: outsider disclosures are restricted [F12] and also that the defendant used these restricted materials [F14].

Student So that would be more like ... limited disclosure would be distinguished from general disclosure, or, competitors knowing ...

Tutor So did Data General discredit your theory?

Student Not in a wide theoretical ... only on the basis of these factors but not in a theoretical sense

Figure 3: Using CATO to test a proposed legal theory

5. Conclusions

In an empirical experiment, we found that human-guided instruction with the CATO environment is as effective as classroom instruction that teaches the same material. Given that arguments and explanations of cases' relevance generated by the CATO program were graded better than any student's answers, it appears that the CATO model can be employed even more effectively. We continue to investigate how instruction with CATO can be improved, in particular by making explicit CATO's argument plan and relevance criteria, by providing more examples of arguments, and by providing more top-down guidance in constructing arguments.

Examples from protocols of human-lead CATO instruction indicate that students were able to use the CATO con-

ceptual framework and tools effectively. Consistently with the model's intent, they used factors to (a) represent the strengths and weaknesses in a problem, (b) make initial assessments of whether cases are relevant, and (c) organize and express arguments. But the students also used factors in ways that went beyond the model, to: (d) implement general argumentation/research strategies, (e) formulate issues and focus research, and (f) formulate general theories about trade secrets law and retrieve cases to test them.

Based on these data and examples, we conclude that the CATO model and instructional environment guide and facilitate legal research, and provide useful opportunities for practice, at least when students are guided by a human tutor. Providing factors helps to get students started in comparing and contrasting cases and making arguments. The model and case representations based on the model reduce some of the distracting complexity and allow students to complete a few cycles of the legal research task.

Even though the model does not represent certain aspects of reasoning with cases and factors, the evaluation results show that students can learn useful knowledge and skills, and the examples show that students are using factors in ways that go beyond the model's limitations. In order to help students deal with the model's limitations, we provided squibs, narrative summaries of the cases, and gave students a lecture about trade secrets law prior to their CATO sessions. With these measures in place, the limitations of the model's focus on factors do not appear to have mislead students into reasoning too mechanically with factors. Moreover, the kind of reasoning CATO does engage in appears to prompt students to do interesting things with factors.

References

Aleven, V., and K. D. Ashley, 1994. An Instructional Environment for Practicing Argumentation Skills. In AAAI-94: *Proceedings of the Twelfth National Conference on Artificial Intelligence,* 485–492. Menlo Park, CA: AAAI Press.

Aleven, V., and K. D. Ashley, 1993. What Law Students Need to Know to WIN. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Law,* 152–161. New York: ACM.

Ashley, K. D., 1990. *Modeling Legal Argument: Reasoning with Cases and Hypotheticals.* Cambridge, MA: MIT Press.

Ashley, K. D., 1991. Reasoning with Cases and Hypotheticals in HYPO. *International Journal of Man-Machine Studies,* 34, no. 6: 753–796.

Ashley, K. D., and V. Aleven, 1994. A Logical Representation for Relevance Criteria. In *Topics in Case-Based Reasoning; Selected Papers from the First European Workshop, EWCBR-93,* edited by S. Wess, K. D. Althoff, and M. M. Richter, 338–352. Lecture Notes in Artificial Intelligence, 837. Berlin: Springer-Verlag.

Ashley, K. D., and V. Aleven, 1992. Generating Dialectical Examples Automatically. In *AAAI-92: Proceedings of the Tenth National Conference on Artificial Intelligence,* 654–660. Menlo Park, CA: AAAI Press.

Berman, D. H., 1991. Developer's Choice in the Legal Domain: The Sisyphean Journey with CBR or Down Hill with Rules. A Working Paper for the Case-Rules Panel. In *Proceedings of the Third International Conference on Artificial Intelligence and Law,* 307–309. New York: ACM.

Berman, D. H., and C. D. Hafner, 1993. Representing Teleological Structure in Case-Based Legal Reasoning: The Missing Link. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Law,* 50–59. New York: ACM.

Clark, P., 1990. Representing Arguments as Background for the Justification of Case-Based Inferences. Turing Institute Research Memoranda TIRM-90-040. Glasgow, Turing Institute Press.

Croft, W. B., and H. R. Turtle, 1992. Text Retrieval and Inference. In *Text-Based Intelligent Systems,* edited by P. Jacobs, 127–155. Hillsdale, NJ: Lawrence Erlbaum.

Hage, J., 1993. Monological Reason-Based Logic; A Low-Level Integration of Rule-Based Reasoning and Case-Based Reasoning. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Law,* 30–39. New York: ACM.

Loui, R. P., J. Norman, J. Olson, and A. Merrill, 1993. A Design for Reasoning with Policies, Precedents, and Rationales. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Law,* 202–211. New York: ACM.

McCarty, L. T., 1994. Defeasible Legal Rules. In Working Notes, AAAI Workshop on Computational Dialectics, edited by R. P. Loui and T. Gordon. Sponsored by AAAI. Murbach, R., and E. Nonn, 1993. Similarity in Harder Cases. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Law,* 236–244. New York: ACM.

Rissland, E. L., D. B. Skalak, and M. T. Friedman, 1993. BankXX: A Program to Generate Argument through Case-Base Search. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Law,* 117–124. New York: ACM.

Rissland, E. L., D. B. Skalak, and M. T. Friedman, 1994. Heuristic Harvesting of Information for Case-Based Argument. In AAAI-94: *Proceedings of the Twelfth National Conference on Artificial Intelligence,* 36–43. Menlo Park, CA: AAAI Press.

Sanders, K. E., 1991. Representing and Reasoning about Open-Textured Predicates In *Proceedings of the Third International Conference on Artificial Intelligence and Law,* 137–144. New York: ACM.

Shute, V. J., and J. W. Regian, 1993. Principles for Evaluating Intelligent Tutoring Systems. *Journal of Artificial Intelligence and Education,* 4, no. 2/3, 245–271.