



Using background knowledge in case-based legal reasoning: A computational model and an intelligent learning environment

Vincent Aleven

*Human–Computer Interaction Institute, Carnegie Mellon University, 5000 Forbes Ave,
Pittsburgh, PA 15213, USA*

Received 7 September 2001

Abstract

Researchers in the field of AI and Law have developed a number of computational models of the arguments that skilled attorneys make based on past cases. However, these models have not accounted for the ways that attorneys use middle-level normative background knowledge (1) to organize multi-case arguments, (2) to reason about the significance of differences between cases, and (3) to assess the relevance of precedent cases to a given problem situation. We present a novel model, that accounts for these argumentation phenomena. An evaluation study showed that arguments about the significance of distinctions based on this model help predict the outcome of cases in the area of trade secrets law, confirming the quality of these arguments. The model forms the basis of an intelligent learning environment called CATO, which was designed to help beginning law students acquire basic argumentation skills. CATO uses the model for a number of purposes, including the dynamic generation of argumentation examples. In a second evaluation study, carried out in the context of an actual legal writing course, we compared instruction with CATO against the best traditional legal writing instruction. The results indicate that CATO's example-based instructional approach is effective in teaching basic argumentation skills. However, a more "integrated" approach appears to be needed if students are to achieve better transfer of these skills to more complex contexts. CATO's argumentation model and instructional environment are a contribution to the research fields of AI and Law, Case-Based Reasoning, and AI and Education.

© 2003 Elsevier B.V. All rights reserved.

Keywords: AI and Law; Case-based reasoning; Similarity assessment; Modeling argumentation; Intelligent learning environments; Intelligent tutoring systems

E-mail address: aleven@cs.cmu.edu (V. Aleven).

1. Introduction

Researchers in the field of AI and Law have long been interested in developing computational models of case-based legal reasoning. While one may think of the law as a system of rules, attorneys very frequently make arguments based on past cases [45]. “[C]ases are the grist of the legal reasoning mill” [24, p. 11].¹ A number of AI and Law models have captured major aspects of the arguments that attorneys make based on past cases. The HYPO model focused on the comparing and contrasting of cases in terms of “Dimensions” in order to generate arguments and explore hypothetical modifications of a problem [7,8]. GREBE [21] and CABARET [54] gave accounts of the subtle ways in which skilled attorneys combine reasoning with cases and with statutory rules. BANKXX modeled the use of case-based reasoning in the context of a search-based approach to argument construction [55]. These models help to clarify and test hypotheses about processes of reasoning with cases in the legal domain. They also provide a potential basis on which to build software applications.

In this paper, we present results from a project investigating whether and how a computational model of expert reasoning with cases can be used to support computer-based instruction for beginning law students. In order to support effective instruction, a model of case-based reasoning must address the types of case-based arguments on which previous models have focused, such as analogizing and distinguishing cases, citing cases as counterexamples, and selecting the best cases to cite. But it must address additional skills that legal experts have and that beginning law students must learn, most importantly, organizing a written argument based on multiple cases and making arguments about the similarity of cases. These case-based argumentation phenomena have so far received little attention in the AI and Law and case-based reasoning (CBR) literature.

The primary claim of this paper is that in order to address these argumentation phenomena, it is necessary to represent and apply *middle-level normative background knowledge*. Constructing such a more complete model presents several challenges. First, the normative background knowledge must be represented in a suitable format. Second, methods must be devised that detail how this knowledge can be applied to generate the various types of arguments mentioned above. With respect to the ways attorneys organize arguments by issues, the challenge is to produce arguments focused on the main issues that a given problem scenario brings up, while marshaling the available cases effectively within this overall structure. The challenge in modeling arguments about the significance of differences between cases is to apply the background knowledge in a context-sensitive manner [12,13,40], so that the significance of a given distinguishing factor may vary depending on the specifics of the cases being compared and the purpose for which the argument is made. We present an argumentation model that, we claim, meets these

¹ Attorneys reason with cases for two main reasons. In some areas of the law, called “common law” areas, the legislature has not provided a detailed statute. The law in these areas has developed mostly through court decisions and cases are thus the primary sources of law. Examples of such areas are tort law and contract law. Even in areas where the legislature has provided a statute, the statute often contains open-textured terms, that is, terms for which no definition is given but that are too abstract to be applied directly to the facts of a case. Past cases provide information about how courts have interpreted open-textured terms.

challenges. It includes a representation of middle-level normative background knowledge for one area of the law, trade secrets law, in the form of a *Factor Hierarchy*. Further, it contains computational methods for using the background knowledge to (1) generate multi-case arguments organized by issues, (2) make context-sensitive arguments about the significance of differences between cases, and (3) select the best cases to cite in an argument. The model has been implemented in the CATO program. To support the claim that normative middle-level background knowledge is an important ingredient of case-based legal argument, we present results from an evaluation study that focused on the question whether the use of the background knowledge may be helpful in predicting the outcome of trade secrets cases. The claim is relevant to the research fields of AI and Law as well as CBR.

A second main claim of the paper is that the argumentation model outlined above can be used to support effective computer-based instruction. More specifically, the claim is that basic argumentation skills can be taught effectively by an intelligent learning environment that provides instruction primarily by presenting argumentation examples. This claim is relevant to the research fields of AI and Law as well as AI and Education. We present the CATO intelligent learning environment, which supports practice of theory-testing and case-based argumentation tasks. CATO uses the argumentation model described above to generate argumentation examples dynamically, to make argument structure visible, and to reduce some of the distracting complexity that makes legal research and analysis tasks hard for beginning students. To support the claim that such an approach can be effective, we present results from an evaluation study, carried out at the University of Pittsburgh School of Law, in which we compared the effectiveness of instruction with CATO against that of more typical legal writing instruction.

The paper is structured as follows: First, we discuss the role of background knowledge in case-based legal reasoning and discuss relevant related research. Next, we present and illustrate CATO's argumentation model, focusing on its main innovations: a representation of middle-level normative background knowledge and methods for using the background knowledge to make arguments with cases. We present results from a study in which we evaluated how well the novel methods for reasoning about the significance of differences between cases can be used to predict the outcome of trade secrets cases, compared to other CBR methods and inductive learning algorithms. We go on to describe the CATO intelligent learning environment and the evaluation study comparing the pedagogical effectiveness of instruction with CATO against that of more traditional legal writing instruction. We conclude by discussing the implications of these evaluations as well as the contributions that the work makes to the research fields of AI and Law, CBR, and AI and Education.

2. Related research in the fields of AI and Law and of CBR

In the current section, we outline the role that middle-level normative background knowledge plays in case-based legal argumentation. We discuss why modeling the use of this kind of knowledge is a contribution to the research fields of AI and Law and CBR.

Although it has been noted that factor-based models of case-based legal argumentation, such as that employed in HYPO [7,8], lack the knowledge to generate certain advanced types of legal arguments, those critiques have focused on high-level knowledge about the policies and purposes underlying the domain [16,17] or knowledge about how the procedural setting of cases influences their value as precedents [18]. The current work focuses on the role of a related, but different kind of background knowledge, namely, middle-level normative knowledge about the meaning of the factors that are used to represent cases.

An important skill that beginning law students must learn is to organize written arguments by issues [50]. Attorneys very frequently organize arguments by issues. Courts explain their decisions in case opinions organized by issues. In the legal domain, as elsewhere, an issue is a point of contention among arguers. In legal disputes, issues arise in many different ways. There may be issues as to what the relevant law is, what the facts of a case are, and how the law should be applied to the facts of a case. Such issues come from a variety of sources, such as statements of the elements of a claim found in a statute or restatement, or tests that courts use, which are often found in cases.

Previous research in the field of AI and Law has focused on modeling the processes that are involved in identifying issues in a legal problem. As Gardner [35] has pointed out, in order to identify issues, a program must discriminate between what is controversial and what is obvious and must focus its attention accordingly. Being able to do so is a hallmark of intelligent argumentation and of intelligence, generally. Previous programs focused on “issue-spotting” in the domain of contract law [35] and tort law [30], using a variety of knowledge sources. The current work focuses on how to use cases effectively in arguments organized by issues, in support of a conclusion that one side or the other (plaintiff or defendant) should win in a given dispute. As they construct such arguments, skilled attorneys identify issues that a problem raises, judge the relevance of available past cases with respect to these issues, and marshal the cases in a variety of argument moves in order to address a party’s strengths and weaknesses with respect to each issue. In each of those steps, they draw on middle-level normative background knowledge. We present a *method for organizing Issue-Based Arguments* which details how skilled attorneys use background knowledge as they construct arguments supported by cites to multiple cases. The method for identifying issues presented here may not be as elaborate or selective as those of, for example, Gardner’s program, but the method provides a more comprehensive model of the ways attorneys marshal cases effectively in an overall argument organized by issues. This method is not specific to any particular area of the law.

A second important skill that beginning law students must learn is to interpret and make arguments about the similarity of cases. In adversarial domains such as the law, opposing parties typically offer competing analogies to a problem. Whether arguers succeed or fail depends critically on their ability to convince a decision maker (a judge or court of law) that the precedents they have presented are more relevant than those relied upon by their opponent. But how to produce such arguments? More often than not, when comparing a problem and a past case, one sees similarities as well as differences. The significance of similarities and differences depends on context [13] and should be interpreted “according to the law” [24, p. 31ff]. But there is no generally-accepted, detailed

method for weighing the similarities and differences. And “it is most difficult to give a satisfactory account of what this [that is, the weighing of similarities and differences, in a context-sensitive manner, and according to the law] might mean in common law adjudication” [24, p. 31].

Nonetheless, attorneys frequently make arguments in which they assess, characterize, and explain the significance of similarities and differences between cases in a context-sensitive manner and in accordance with the law. They do so by drawing on normative background knowledge, interpreting the meaning of the similarities and differences in more abstract legal terms. An important aspect of these arguments is the strategic choice of abstract terms with which to describe and interpret cases. Legal cases can naturally be described at multiple levels of abstraction. These abstraction levels provide an important degree of freedom in generating alternative interpretations of cases [44,45]. Legal professionals are adept at exploiting this freedom in making arguments with cases, characterizing cases as instances of one abstract term or another.

If a computational model is to provide an adequate account of how experts use cases in the legal domain, they must model the strategic interpretation of cases in support of a conclusion that they are similar or not. Previous models however have not addressed this type of argumentation. By and large, they lack the knowledge needed to reason about the significance of similarities and differences and do not have methods to characterize and re-characterize cases in order to argue that they are similar or different. The model presented in the current paper has both. Its *Factor Hierarchy* provides the normative background knowledge needed to reason about the meaning of differences among cases. Further, the model contains *heuristic criteria for selecting focal abstractions* that specify how to employ the background knowledge strategically, to characterize and re-characterize cases, in arguments that distinctions matter or not. These criteria are not specific to any area of the law. Rather, they capture general ways of making arguments about the significance of differences between cases. They achieve the context-sensitivity called for by Ashley and Rissland [13].

In order to produce the strongest possible argument that a given problem should be decided in a favorable manner, an arguer must not only be able to use available cases effectively, but must also be able to select the best cases to cite. Existing interpretive CBR programs have demonstrated various interesting methods for selecting the most relevant cases. The HYPO program used a most-on-point criterion to find cases with maximum sets of shared factors, as well as methods to explore how certain hypothetical modifications of the problem cause the balance of an argument to shift [7,8]. GREBE used a combination of A* search and structure mapping to evaluate the similarity of cases [21]. For the most part, these methods have not placed much emphasis on evaluating the significance of distinctions, Branting’s methods for match improvement perhaps coming the closest. From a rhetorical point of view, however, it makes sense that the significance of distinctions should be taken into account when selecting the best cases to cite. After drawing an analogy between a problem and a past case, an arguer would strongly prefer not to see the opponent point out that there are significant distinctions. We present a number of novel relevance criteria that focus on cases that do not have significant distinctions.

Accurate judgment of the similarity of cases is a central concern in CBR research [40,42]. Smyth and Keane [65] called for deeper methods for similarity assessment and

demonstrated one approach suitable primarily for problem-solving CBR. In the current paper, we focus on deeper similarity assessment in interpretive CBR, where it is no less important than in problem-solving CBR. The arguments about the significance of distinctions modeled in CATO are a dialectic way to achieve deeper similarity assessment, in which the similarity of cases is evaluated at multiple levels of abstraction and from different viewpoints.

Within the field of CBR, a number of advanced techniques have been developed for reasoning about and explaining aspects of similarity, but none of these methods quite fit the constraints of adversarial, interpretive CBR. CASEY reasons about the significance of differences using a causal model of heart disease that by itself is sufficient to solve problems “from scratch” [41]. Protos, a case-based reasoning program that was applied to a number of domains including auditory disorders, engages in knowledge-based explanation of the similarity between cases [14]. TRUTH-TELLER assesses and explains the similarity or dissimilarity of truth-telling dilemmas using hierarchical background knowledge that in itself is insufficient to solve problems [46]. ACCEPTER’s abstraction nets [43] are designed to represent general patterns of explanation and support a wide range of alternative interpretations of an event (such as the death of a race horse). However, none of these programs place much emphasis on generating alternative interpretations of a distinction’s significance, as does CATO.

The methods for reasoning about the significance of distinctions among cases presented in this paper relate to a second important theme in CBR research, namely, abstraction. CATO’s methods are designed to model the strategic selection of abstract interpretations of cases in order to support an argument that they are similar or that they are different. Abstraction has been used in many phases of the CBR process, including retrieval, similarity assessment, and adaptation [15,47,69]. The use of abstraction in CBR approaches to problem solving and planning, often referred to as “stratified CBR”, has been shown to lead to considerable improvements in efficiency [15,19,22]. But stratified CBR relies on the ability to define abstraction layers, transform problems with respect to these layers, and deal with one abstraction level at a time. In the legal domain, one rarely finds such neatly-ordered layers of abstraction. Much more often, attorneys must simultaneously oversee multiple levels of abstraction, taking into account that conflicting evidence may exist for abstractions at different levels. Thus, in case-based legal reasoning one faces some of the same problems related to the use of abstract terms as are encountered in conversational CBR [36], although in the legal domain they are compounded by the adversarial nature of the domain.

3. A representation of middle-level normative background knowledge

In the current and the next three sections, we describe and illustrate the novel aspects of CATO’s argumentation model. We start with the representation of normative background knowledge in the Factor Hierarchy. In the next sections we describe and illustrate how this representation of background knowledge is key to modeling a number of argument types not previously addressed in AI models of legal reasoning.

A representation of middle-level, normative background knowledge

A *Factor Hierarchy* represents domain-specific normative knowledge about the meaning of the factors used to represent cases. Factors were first used in the HYPO program [7,8].

Methods for generating arguments that use background knowledge

Issue-Based Arguments are multi-case arguments organized by issues; they involve the use of middle-level normative background knowledge and the use of cases in various argument moves.

Arguments about the significance of distinctions emphasize or downplay the significance of a given distinction in the comparison of a problem and a precedent case.

Methods for evaluating the relevance of past cases that use background knowledge

There are two main criteria for selecting the best case to cite in an argument, both of which involve the evaluation of the significance of distinctions. The criteria build on HYPO's most-on-point criterion [7,8].

Fig. 1. Novel components of CATO's model of case-based legal argumentation.

The CATO model provides an account of arguments in which two parties argue for a favorable outcome in a given problem situation, comparing and contrasting the problem to relevant past cases. The main novel components of the model are listed in Fig. 1. The model also includes elements from the HYPO model of case-based legal argumentation developed by Ashley [7,8], such as the use of factors to represent cases and a number of basic argument moves, such as the analogizing and distinguishing of cases in terms of factors.²

We illustrate CATO's argumentation model in the context of a trade secrets problem based on the Missouri case *National Rejectors, Inc., v. Trieman*, 409 S.W.2d 1 (Mo.1966). *National Rejectors v. Trieman* is a dispute between a manufacturer of coin-handling devices and a number of its former employees who had started a competing business. National Rejectors, Inc. complained that its former employees had misappropriated its designs for coin-handling devices, which it claimed were trade secrets. A short summary of the facts of the case is shown in Fig. 2, on the left. Trade secrets law aims to protect owners of commercially valuable information against unfair competition by parties who obtained that information through a breach of confidence or through improper means, provided that the information qualifies as a trade secret. The parties in *National Rejectors* (plaintiff and defendant) must make the strongest possible argument that they should win a claim of trade secrets misappropriation. CATO's argumentation model provides an account of how past cases can be used to support such arguments.

3.1. Background: Use of factors to represent cases

Before describing the Factor Hierarchy, we discuss how cases are represented. In CATO, cases are represented in terms of *factors* [7,8]. Factors are stereotypical collections of facts

² HYPO, which was an argumentation program rather than an instructional program, had a number of capabilities that were not adopted in CATO. For example, the "Dimensions" used to represent cases in HYPO capture more information than CATO's factors [8,53,56]. Also, HYPO was able to identify "hypothetical modifications" of a problem situation that led to a shift in the balance of an argument. The fact that these capabilities were not included in the CATO program does not reflect a judgment about their potential utility for computer-based instruction.

Facts	Factors
National Rejectors, Inc., v. Trieman	
Since the 1940s, <i>National</i> was practically the sole supplier of rejectors and changers, coin-handling devices used in vending machines and washers. National had developed its products through many years of experimentation. In 1957, National employees, defendants Trieman and Melvin, started their own business for producing coin-handling devices.	F15 Unique-Product (p)
<i>Melvin, working at his home, designed two rejectors that were as close as possible to the comparable National rejectors. He combined his knowledge of the National device with information obtained from measuring National rejectors. He also used production drawings, a few parts, and materials obtained, without consent, from National.</i>	F18 Identical-Products (p)
<i>However, none of defendants' drawings was shown to be a copy of a drawing of National's. The resulting rejector improved on the National product in certain ways.</i>	F7 Brought-Tools (p)
Melvin and Trieman resigned from National in March and July, 1959, respectively. Twelve other National employees went to work for Coin Acceptors, after initiating contact themselves. One of them used information remembered from National's drawings. In a two-year period, Coin Acceptors manufactured about 150,000 single-coin units.	
<i>National's vice-president testified that a skilled mechanic could make drawings by taking apart the National rejectors and measuring the parts. Information about the parts was publicized in National's patents as well as in catalogs, brochures, and manuals distributed to customers and to the trade generally. National did not take any steps at its plant to keep its information confidential. Employees were not required to sign non-competition agreements or notified of confidentiality. Engineering drawings were sent to customers and prospective bidders without limitations on their use.</i>	F16 Info-Reverse-Engineerable (d)
	F27 Disclosure-In-Public-Forum (d)
	F19 No-Security-Measures (d)
	F10 Secrets-Disclosed-Outsiders (d)

Fig. 2. Textual description of case facts and the applicable factors.

that, experts agree, influence the outcome of a case. Although generally a factor is neither necessary nor sufficient to decide a legal claim, all else being equal the presence of a factor makes a case stronger or weaker for the plaintiff. CATO has a set of 26 factors for the domain of trade secrets law, 12 of which were first used in HYPO. Some of these factors were gleaned from the list of factors presented in the Restatement of Torts §757, comment b, which until recently many courts have adopted as an authoritative statement of the law of trade secrets. Other factors were extracted from cases and secondary sources.

For example, in the *National Rejectors* problem shown in Fig. 2, the plaintiff's position is strengthened by the fact that its product was unique on the market (factor F15, Unique-Product). Also, one of the defendants took product development information and tools from the plaintiff's business (F7, Brought-Tools) and the defendant manufactured a product that was highly similar to the plaintiff's (F18, Identical-Product). On the other hand, in the defendant's favor, the plaintiff did not take any measures to protect the secrecy of its information (F19, No-Security-Measures) and disclosed information to outsiders (F10, Secrets-Disclosed-Outsiders) and even to the trade and general public (F27, Disclosure-In-Public-Forum). Also, the information that the plaintiff claimed to be its trade secret

could be learned by taking apart and examining the plaintiff's devices (F16, Info-Reverse-Engineerable).³

Given that some of the factors that apply in the *National Rejectors* problem favor the plaintiff, while others favor the defendant, who should win? In law, there is no authoritative weighting scheme that can be used to decide whether the pro-plaintiff factors outweigh the pro-defendant factors, or vice versa [8]. Instead, arguers make arguments based on past cases, which they typically organize by the issues that are raised by the given problem situation. The construction of such arguments involves the use of background knowledge about the meaning of the factors used to represent cases.

3.2. The Factor Hierarchy

One of CATO's innovative features is its representation of middle-level normative knowledge, called the Factor Hierarchy, part of which is shown in Fig. 3. The Factor Hierarchy is not a rendition of an existing authoritative or established conceptual hierarchy for the domain of trade secrets law, but was constructed following a knowledge engineering approach. The Factor Hierarchy covers the basic requirements of a claim for trade secret misappropriation.⁴ The Factor Hierarchy is not meant to state necessary or sufficient conditions for winning a claim of trade secrets misappropriation. The knowledge represented in the Factor Hierarchy is weaker than that. In this sense, CATO's Factor Hierarchy is different from Branting's proposed *reduction graphs* for representing the *ratio decidendi* of cases [20], which are intended to represent legal knowledge at multiple levels of abstraction, just as is done in the Factor Hierarchy.

In the Factor Hierarchy, the 26 factors used to represent cases, some of which are shown at the bottom in Fig. 3, are linked to normative concepts: 11 so-called Intermediate Legal Concerns in the middle tiers, which in turn are linked to 5 Legal Issues at the top of the Factor Hierarchy. We use the terms "abstract factors" or "high-level factors" to refer to the Intermediate Legal Concerns and Legal Issues. Each high-level factor stands for two opposing conclusions, one favoring the plaintiff, one favoring the defendant. For example, associated with abstract factor F102, Efforts-To-Maintain-Secrecy, are a pro-plaintiff conclusion that "Plaintiff took efforts to maintain the secrecy of its information"

³ In order to add a case to CATO's Case Database, or to input a new problem situation, it must first be represented in terms of factors. To ascertain which factors apply, the case entry person must read the opinion, the document in which the court summarizes the facts of the case and explains its decision. Occasionally, this process leads to the discovery of new factors. If the case is to be used in CATO's instructional environment, the case entry person must also create a "squib", a one page summary of the court's opinion. Adding a new case typically takes several hours. It has been our experience that even with experienced case entry people there remains an element of subjectivity in deciding which factors apply in a given case.

⁴ To give a more specific idea of how much ground the Factor Hierarchy covers: In an American law school, the material in the Factor Hierarchy represents about one fifth of a three credit course on intellectual property law, an increasingly popular elective course in a normal three-year, eighty-eight credit course of study. We stress, however, that CATO's instructional environment was designed to teach argumentation skills to beginning law students, not trade secret law. Its argumentation model is not specific to trade secrets law, even if its Factor Hierarchy and case database are.

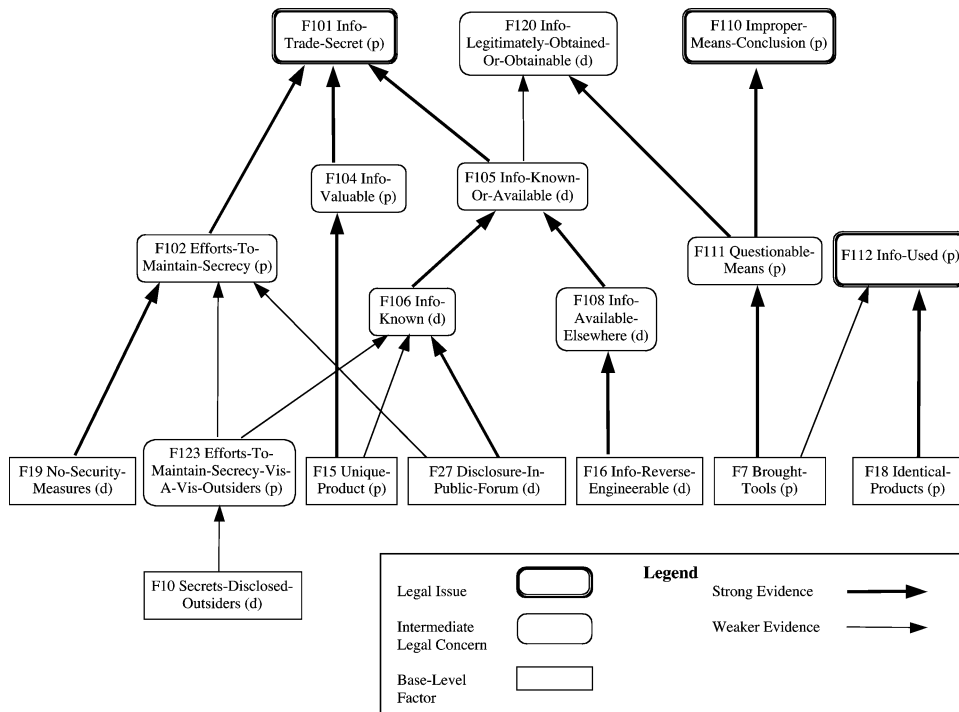


Fig. 3. Excerpts from CATO's factor hierarchy.

and a pro-defendant conclusion that “Plaintiff showed a lack of interest in maintaining the secrecy of its information”.

A positive link in the hierarchy indicates that a factor lends (defeasible) support to a more abstract factor, in other words, that it tends to favor a conclusion for a particular side with respect to a normative concept. A negative link indicates that a factor tends to favor an opposite conclusion. In Fig. 3, for example, all of the factors and abstract factors linked to F102, Efforts-To-Maintain-Secrecy, provide evidence for or against the conclusions associated with F102, which itself provides evidence concerning the more abstract issue of whether the information is a trade secret, F101, Info-Trade-Secret. Links can be strong (thick links) or weak (thin links), indicating the level of support that they represent. CATO takes into account the link strength in order to evaluate the support for abstract factors, as is discussed further below. The links are assumed to be self-evident, grounded in the common sense of the legal claim. Roth [57] recently proposed ways of modeling arguments about links in Factor Hierarchies such as CATO's, which have the potential to lead to interesting extensions of CATO's current argument moves.

The Legal Issues at the top of the Factor Hierarchy correspond to the main issues that courts often address in explaining their decisions. Some of the Legal Issues are gleaned from the Restatement 1st of Torts, §757, which provides that one is liable for misappropriating another's trade secret if one uses that secret in breach of a confidential

relationship or uses improper means to acquire the secret. The corresponding Legal Issues are F114, Confidential-Relationship and F110, Improper-Means-Conclusion. Other issues come from cases or secondary literature. Intermediate Legal Concerns correspond to concepts used by legal experts to analyze cases and explain decisions in cases, but they are subordinate to the Legal Issues. The Intermediate Legal concerns related to the issue F101, Info-Trade-Secret correspond to factors listed in the Restatement 1st of Torts, §757. Others come primarily from cases.

Although we illustrate CATO's argumentation model in the domain of trade secrets law, the model itself is not specific to that area of the law. Factor representations of cases have been developed for tax law in the CABARET system [54] and bankruptcy law in BANKXX [55]. Also, we have developed a CATO case database with cases about the employee/independent distinction and have identified factors for the fair use concept in copyright law. No Factor Hierarchies were developed for any of these areas. Rissland [52] has shown that some of the arguments made before the United States Supreme Court involve hypothetically modifying cases by moving them along "Dimensions", of which CATO's factors are slightly simplified versions.

Based on our experience in the area of trade secrets law, we estimate that it would take 2–4 months to develop an initial case database and Factor Hierarchy for a new area of the law, depending on the knowledge engineering experience of the developers, their degree of legal expertise, and the level of involvement of legal experts. Ideally, one would interact regularly with legal experts, asking them to make arguments with carefully selected sets of cases, including arguments about the significance of similarities and differences. Citation services such as those offered by the Westlaw™ legal information retrieval system, which lists for any given case the cases that have distinguished it, are likely to be very helpful in selecting small groups of cases that make interesting comparisons.

4. Using background knowledge to organize multi-case arguments by issues

Having described a representation of background knowledge, we now present a model that details how the background knowledge can be used for purposes of argumentation. A second innovation of the CATO model is a computational method for generating multi-case arguments organized by issues. In this section, we describe this method, highlighting how CATO uses the background knowledge represented in its Factor Hierarchy to identify issues in a problem and to marshal cases.

4.1. Identifying issues

The first step in generating Issue-Based Arguments is to identify the main issues that a problem raises. The CATO model focuses on issues of how the law should be applied to the facts of a case, related to some of the important concepts in trade secrets law, such as those found in the Restatement 1st of Torts. Which issues are raised in a particular problem depends on which factors are present. CATO raises an issue when a problem or case presents evidence related to it in the form of base-level factors, regardless of which side these factors favor. That is, to identify issues in a problem or case, CATO

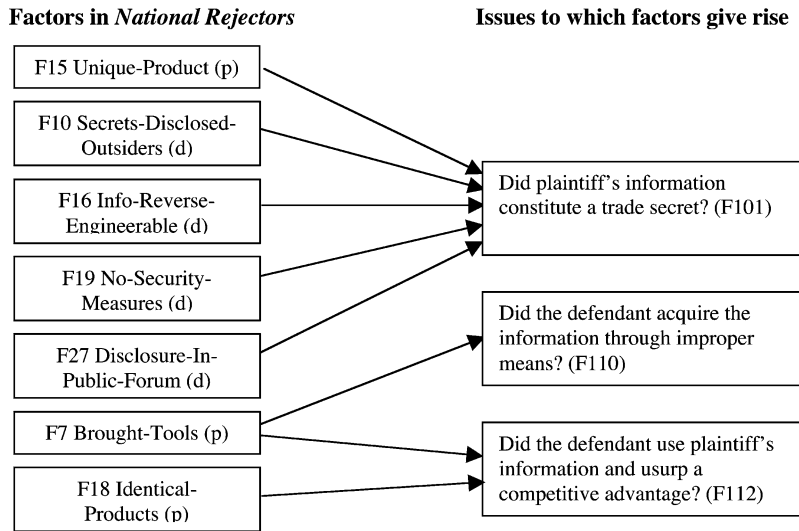


Fig. 4. CATO identifies issues related to the applicable factors of a case.

collects all Legal Issues in the Factor Hierarchy that are linked to the case's applicable factors.

Following this method, CATO identifies three issues raised by the *National Rejectors* factors, shown in Fig. 4. There are conflicting factors with respect to the issue of whether the plaintiff's information is a trade secret (F101). For example, the plaintiff can cite the fact that its product was unique in the market (pro-plaintiff factor F15) as evidence that its information should be considered a trade secret. The defendant, on the other hand, could point to the disclosures that the plaintiff made to outsiders (F10) and the general public (F27) as evidence that supports the opposite conclusion. It appears that the plaintiff's position is rather weak with respect to this issue, since many of the related factors favor the defendant, but one cannot really know without looking at past cases. There are no conflicting factors related to the other two issues. The plaintiff's position with respect to these issues appears to be much stronger than for the first issue, but there is still the question whether the pro-plaintiff factors provide sufficient evidence to warrant a conclusion favorable to the plaintiff with respect to those issues.

Thus, like other programs, CATO raises an issue when there is uncertainty about a conclusion. Gardner's system for the domain of contract law raised issues when it detected that it had either had insufficient knowledge or conflicting knowledge related to a question [35]. CHASER, a system for tort law, raises issues when it finds arguments for both sides related to an element of a tort claim [30]. CATO is perhaps less selective or specific when raising issues than these programs, but it is unique in that it uses the same representation of background knowledge to organize arguments by issues and to reason about the significance of differences of cases.

4.2. Organizing arguments by issues

Having identified issues, the next step is to generate an argument that addresses each issue, as illustrated in Fig. 5, which shows an Issue-Based Argument generated by CATO on behalf of the defendant in *National Rejectors*. During the instruction with CATO, these arguments serve as models for the (written) arguments that students are asked to make, as discussed in a later section of the paper.

CATO's argument about each issue follows a basic rhetorical design, shown in the annotations on the right in Fig. 5. CATO first emphasizes the strengths related to the issue and then counteracts any weaknesses that may be present, using cases to support its points. Further, when arguing on behalf of the defendant, CATO responds to points made by the plaintiff in its argument (not shown). CATO uses four basic argument moves to employ cases within its basic rhetorical strategy. When making an affirmative argument, it cites cases to emphasize strengths and downplay weaknesses related to an issue. When responding to an argument made by the opponent, it distinguishes cases and cites cases as counterexamples. The latter two argument moves were first used in the HYPO program [7,8]. In constructing arguments focused on issues, CATO uses the background knowledge represented in its Factor Hierarchy extensively, namely,

- to identify issues in a problem;
- in the discussion of an issue, to focus on the factual strengths and weaknesses (factors) that are related to the issue and to cite cases to emphasize strengths and downplay weaknesses;
- to give reasons why particular factual strengths matter to an issue being discussed;
- to find strengths that are closely related to weaknesses and therefore compensate for those weaknesses.

For example, in its argument about the issue of whether the plaintiff's information is a trade secret, represented in the Factor Hierarchy by Legal Issue F101, Info-Trade-Secret, CATO cites the defendant's strengths related to this issue, namely, the fact that the plaintiff disclosed its purported trade secret in various ways (factors F10 and F27) without taking any measures to keep the information secret (F19) and the fact that the information could be learned by reverse-engineering the plaintiff's product (F16). The strengths correspond to the pro-defendant factors present in *National Rejectors* that are linked to the Legal Issue F101 in the Factor Hierarchy (see Fig. 3). References to factors and abstract factors are shown in Fig. 5 in square brackets. To emphasize these strengths, CATO characterizes their meaning in more abstract terms: "plaintiff showed a lack of interest in maintaining . . . secrecy" and "plaintiff's information was known in the industry or available from sources outside plaintiff's business". These phrases correspond to Intermediate Legal Concerns selected from the Factor Hierarchy. Specifically, CATO selects the Intermediate Legal Concerns that connect the factors and issue of interest, F101 Info-Trade-Secret, namely, F102, Efforts-To-Maintain-Secrecy, and F105 Info-Known-Or-Available (see Fig. 3). In order to further emphasize the defendant's strengths, CATO cites two cases in which similar sets of strengths were present and the defendant won: *Yokana* and *Dynamics*. Here,

Argument for Defendant in *National Rejectors, Inc., v. Trieman*

Defendant should win a claim of trade secrets misappropriation. Plaintiff's information is not a trade secret [F101], defendant did not acquire plaintiff's information through improper means [F110], and defendant's information was the result of independent development efforts and investment [F112].

Plaintiff's information is not a trade secret [F101]

Plaintiff's information is not a trade secret [F101]. Restatement 1st of Torts s 757, and Comment b, factors 1–6 (1939). In the problem at hand, plaintiff made disclosures to others outside its business [F10], plaintiff's product information could be learned by reverse-engineering [F16], plaintiff did not adopt any security measures [F19], and plaintiff presented or published the information in a public forum [F27]. This shows that plaintiff showed a lack of interest in maintaining the secrecy of its information [F102] and plaintiff's information was known in the industry or available from sources outside plaintiff's business [F105]. In *Midland-Ross Corp. v. Yokana*, 293 F.2d 411 (3rd Cir.1961) and *Dynamics Research Corp. v. Analytic Sciences Corp.*, 400 N.E.2d 1274, there was similar evidence that plaintiff's information is not a trade secret [F101], and defendant won. *Yokana* held for defendant, where, as in the case at hand, plaintiff made disclosures to others outside its business [F10], plaintiff's product information could be learned by reverse-engineering [F16], and plaintiff made disclosures of the information in a public forum [F27]. In *Dynamics*, as in the current case, plaintiff made disclosures of the information in a public forum [F27], and defendant won.

The fact that plaintiff was the only manufacturer making the product [F15] does not necessarily imply a conclusion that plaintiff's information is a trade secret [F101], especially where plaintiff's information could be ascertained by examining or reverse engineering plaintiff's product [F16], plaintiff made disclosures to others outside its business [F10], and plaintiff presented or published the information in a public forum [F27]. In *Dynamics*, defendant won even though plaintiff was the only manufacturer making the product [F15]. In *Dynamics*, plaintiff presented or published the information in a public forum [F27].

The cases cited by plaintiff with respect to this issue, *Emery*, *Schreyer*, and *Allen*, can be distinguished. Moreover, they are not necessarily the most relevant cases. Defendant can cite a counterexample to some of these cases.

Schreyer is distinguishable. It is stronger for plaintiff than is the current problem. In *Schreyer*, defendant saved time or money by using plaintiff's information. This was not so in *National Rejectors*. Also, in *National Rejectors*, plaintiff disclosed its product information to outsiders and plaintiff did not take any measures to protect the information. This was not so in *Schreyer*. Additionally, *Yokana* is more on point than *Schreyer*, yet held for defendant. In *Yokana*, as in the current problem, plaintiff's former employee took documents, blueprints, or tools to defendant [F7] and plaintiff made disclosures to others outside its business [F10]. This was not so in *Schreyer*. It follows that *Yokana* is more similar to the problem.

[argument distinguishing *Emery* and *Allen*]

Defendant did not acquire plaintiff's information through improper means [F110]

[argument justifying a decision for defendant on this issue and response to cases cited by plaintiff]

Defendant's information was the result of independent development efforts and investment [F112]

To justify a favorable decision on an issue:

Point to strengths, related to issue, say why they matter

Show cases in which these strengths led to a favorable outcome

Discuss weakness(es) related to issue. Point to strengths that compensate.

Show cases that had favorable outcome in spite of weakness(es)

Distinguish cases cited by the opponent and cite counter-examples

Fig. 5. Issue-Based Argument generated by CATO (annotations added).

CATO employs one of its basic argument moves, *Citing a favorable case to emphasize strengths*.

CATO also employs its Factor Hierarchy as it addresses the defendant's sole weakness related to the given issue, the fact that the plaintiff's product was unique in the market (factor F15 Unique-Product). To downplay this weakness, CATO first points to a number of pro-defendant factors present in the *National Rejectors* problem that compensate for it: The plaintiff had made disclosures of the information (F10, F27) and the information could be learned by reverse-engineering the plaintiff's product (F16). To select these *compensating factors*, CATO looks for favorable (that is, pro-defendant) factors that are closely related to the given weakness, meaning that they are linked to an Intermediate Legal Concern that is linked to the given weakness and to the issue of interest. Abstract factor F105 Info-Known-Or-Available qualifies (see Fig. 3). A second way in which CATO downplays the defendant's weakness, F15, is by pointing to a case, *Dynamics*, in which the defendant won in spite of the presence of that same weakness. The point based on *Dynamics* illustrates CATO's basic argument move *Citing a favorable case to argue that weaknesses are not fatal*.

Arguing on behalf of the defendant, CATO also responds to the argument that the plaintiff presumably made. To the extent possible, it distinguishes the cases cited by the plaintiff and cites counterexamples, using the basic argument moves *Distinguishing a case with an unfavorable outcome* and *Citing a more on point counterexample to a case cited by an opponent*. CATO follows the same basic rhetorical approach for the other two issues identified in the problem.

4.3. An algorithm for generating multi-case arguments organized by issues

In order to generate an Issue-Based Argument, CATO must be given a problem situation and a set of precedent cases, all represented in terms factors, and a side on whose behalf to argue (plaintiff or defendant). The set of past cases may include cases won by both sides. If it includes unfavorable cases, CATO will not only make an affirmative argument that the given side should win, citing favorable cases, but will also respond to the points that the opponent could make based on the given unfavorable cases. CATO generates Issue-Based Arguments (written in English) in three steps, shown in Fig. 6: First, using the Factor Hierarchy, it identifies applicable issues and determines which problem strengths and weaknesses are related to each issue. Second, CATO determines for each issue which favorable cases can be used to emphasize strengths or downplay weaknesses related to the issue. It also determines for each case how the opponent could respond, either by distinguishing it or by citing one or more of the given unfavorable cases as counterexamples. If the given set of precedent cases includes unfavorable cases, CATO repeats the same step, this time using the unfavorable cases. This way, CATO determines what points the opponent could make, so that it knows how to respond. Third, using the information collected and organized in the previous two steps, CATO generates the text of the argument, selecting argument templates and filling them out with English phrases stored with each case, factor, or abstract factor.

Generate Issue-Based Argument

Input: a problem situation and a set of past cases represented in terms of factors, and a side on whose behalf to argue (plaintiff or defendant).

Knowledge source: Factor Hierarchy

Output: Multi-case argument organized by issues, written in English

Procedure:

Identify issues

Find all issues related to the problem factors.

For each issue *I*

Determine strengths and weaknesses in the problem related to *I*.

For the strengths, find reasons why they matter (Intermediate Legal Concerns in the Factor Hierarchy).

Organize cases by issues

For each issue *I*

Determine which of the given cases are relevant to *I* (i.e., have factors linked to *I* in the Factor Hierarchy). For each relevant favorable case, determine which of its strengths and weaknesses are related to *I*.

For each weakness related to *I*, check if there are compensating strengths in the problem (i.e., favorable factors that share with the weakness an Intermediate Legal Concern as ancestor). Also, check which of the given cases are can be used to downplay the weakness (i.e., cases with favorable outcome even though they have the weakness).

Check for which of the relevant (favorable) cases there are counterexamples among the given cases (i.e., cases that are more on point or as on point).

Do the same for the opposing side. (When generating an Issue-Based Argument on behalf of the defendant, it is necessary to reconstruct how cases were used in the argument being responded to.)

Generate English text for argument organized by issues

For each issue *I*

Draw attention to strengths related to *I*, pointing to reasons why they matter and citing relevant cases.

Deal with weaknesses related to *I*, pointing to compensating strengths and citing counterexamples.

When arguing on behalf of the defendant, distinguish cases cited by plaintiff when discussing *I* and cite counterexamples.

Fig. 6. CATO's method for generating Issue-Based Arguments.

5. Using background knowledge to reason about the significance of differences

A third innovative feature of the CATO model are its methods for using background to reason about the significance of differences among cases. In legal argumentation, the parties very frequently offer competing analogies to the problem. Much legal argument therefore involves debating whether a case is really the same as the problem or not. One side analogizes the problem to a case, the other side distinguishes the problem. The first side downplays the significance of the differences, the opponent emphasizes them. The dialectical process is one of characterizing and re-characterizing the relevant features of the problem and the cited case in terms of their legal significance in support of an argument either that the cases should be decided alike or differently. Through this dialectic process it usually becomes clear which of the competing analogies is more relevant. While the HYPO program already provided an account of analogizing and distinguishing cases in terms of factors [7,8], CATO is the first computational model of legal argument that focuses on

National Rejectors (d)	Ferranti (d)
* F7 Brought-Tools (p)	F2 Bribe-Employee (p)
F10 Secrets-Disclosed-Outsiders (d)	* F17 Info-Independently-Generated (d)
* F15 Unique-Product (p)	= F19 No-Security-Measures (d)
F16 Info-Reverse-Engineerable (d)	* F20 Info-Known-To-Competitors (d)
* F18 Identical-Products (p)	= F27 Disclosure-In-Public-Forum (d)
= F19 No-Security-Measures (d)	
= F27 Disclosure-In-Public-Forum (d)	= shared factor * distinction

⇒ **Defendant’s argument analogizing the *National Rejectors* problem to the *Ferranti* case**

Where plaintiff did not adopt any security measures [F19] and plaintiff made disclosures of the information in a public forum [F27], defendant should win a claim of trade secrets misappropriation, as in *Ferranti Electric, Inc. v. Harwood*, 43 Misc.2d 533, 251 N.Y.S.2d 612 (1964).

⇐ **Plaintiff’s argument distinguishing the *National Rejectors* problem from the *Ferranti* case**

Ferranti is distinguishable. It is stronger for defendant than is the current problem. In *Ferranti*, defendant developed its product independently of plaintiff’s [F17] and plaintiff’s information was generally known in the industry [F20]. This was not so in *National Rejectors*.

Also, in *National Rejectors*, plaintiff’s former employee removed documents containing plaintiff’s information [F7], plaintiff’s product was different from products made by competitors [F15], and there were substantial similarities between plaintiff’s and defendant’s products [F18]. This was not so in *Ferranti*.

Fig. 7. CATO’s arguments analogizing and distinguishing a problem and a precedent.

arguments about similarity, in which arguers characterize and re-characterize cases in light of abstract background knowledge.

5.1. Background: Comparing and contrasting cases in terms of factors

Before we turn to CATO’s arguments about the significance of distinctions, we first describe how CATO, following the approach pioneered in HYPO [7,8], models arguments analogizing and distinguishing a problem and a past case. Fig. 7 shows a comparison of the factors of the *National Rejectors* problem and *Ferranti*, a case won by the defendant. The two cases have several factors in common, marked with “=” in Fig. 7. CATO, arguing on behalf of the defendant, draws an analogy between the cases, pointing to the shared factors as relevant similarities and claiming that they justify treating the cases alike: *Ferranti* was won by the defendant, so should *National Rejectors*. The referenced factors are shown in square brackets.

While the defendant’s argument is quite convincing, the analogy is not perfect. There are distinctions between the cases, marked with “*” in Fig. 7. CATO arguing on behalf of the plaintiff argues that *Ferranti* is distinguishable and hence is not a relevant precedent, thereby undermining the defendant’s argument. CATO’s argument is shown in Fig. 7. The gist of the argument is that the distinguishing factors are evidence that *Ferranti* is stronger for the defendant than is *National Rejectors*. The extra strengths for the defendant in *Ferranti* and the extra strengths for the plaintiff in *National Rejectors* justify deciding the two cases differently. It is important to note that not all differences are distinctions: CATO’s argument distinguishing the two cases focuses on those unshared factors that help in showing that *Ferranti* is stronger for the defendant

than is *National Rejectors*. Generally, unshared factors are distinctions when they help in showing that the cited case is stronger, for the side that cited it, than the given problem situation. This way of modeling distinguishing was introduced in the HYPO program [7,8].

5.2. Example arguments about the significance of distinctions

Given that *Ferranti* is not a perfect match with *National Rejectors*, how important are the similarities and differences? To assess the significance of similarities and differences of cases, an arguer must relate them to more general (but still domain-specific) normative concepts. In this section we illustrate how CATO models the strategic characterization and re-characterization of cases of which skilled attorneys are capable in order to argue that cases are similar or not. We also illustrate how these arguments satisfy an important desideratum for similarity assessment in CBR, namely, that the significance of distinctions depends on the context in which the argument is made [12,13,40]. In general, schemes for similarity assessment in all but the simplest domains must assign weights to similarities and differences in a context-sensitive manner. The significance of differences must depend on the specifics of the problem and past case being compared and the purpose for which the comparison is made.

To emphasize the significance of a distinction, CATO characterizes the cases as different in the abstract and thus argues that the distinction is indicative of a deeper difference between the cases. Conversely, CATO downplays the significance of a distinction by showing that at a more abstract level, a parallel exists between the cases, arguing in effect that the apparent distinction is merely a mismatch of details. Also, CATO can downplay a distinction by showing an interpretation of the case in which the distinction applies that is opposite to what the presence of the distinction would suggest.

Applying these methods, arguers with opposing viewpoints may offer alternative characterizations of a distinction's significance, as is illustrated in Fig. 8, which shows arguments generated by CATO about the significance of one of the distinctions between *National Rejectors* and *Ferranti*, the fact that in *Ferranti*, but not in *National Rejectors*, the information claimed to be a trade secret was known in the industry (F20). CATO downplays the distinction on behalf of the defendant, interpreting the F20 factor in *Ferranti* as evidence that the plaintiff's information was known outside the plaintiff's business (F105). It points to evidence that warrants the same interpretation of *National Rejectors*: In *National Rejectors*, the plaintiff had disclosed the information to outsiders (factor F10) and even in a public forum (F27). Further, the information could be learned by reverse-engineering plaintiff's product (F16). Thus, the defendant argues, in spite of the apparent distinction, at a deeper level, the cases are similar. The argument illustrates CATO's argument move *drawing an abstract parallel*.

CATO arguing on behalf of the plaintiff is not without retort. It draws attention to an alternative interpretation of the distinction's meaning, one that is equally plausible as the one offered in the argument on behalf of the defendant, and shows that with respect to *that* interpretation, a contrast exists between the cases. In its argument on behalf of the plaintiff, CATO characterizes the F20 distinction as showing that in *Ferranti*, the defendant

National Rejectors (d)	Ferranti (d)
* F7 Brought-Tools (p)	F2 Bribe-Employee (p)
F10 Secrets-Disclosed-Outsiders (d)	* F17 Info-Independently-Generated (d)
* F15 Unique-Product (p)	= F19 No-Security-Measures (d)
F16 Info-Reverse-Engineerable (d)	* F20 Info-Known-To-Competitors (d)
* F18 Identical-Products (p)	= F27 Disclosure-In-Public-Forum (d)
= F19 No-Security-Measures (d)	
= F27 Disclosure-In-Public-Forum (d)	= shared factor * distinction

⇒ Defendant's argument downplaying distinction F20 in *Ferranti*.

In *Ferranti*, plaintiff's information was known to competitors [F20]. This was not so in *National Rejectors*. However, this is not a significant distinction. In *National Rejectors*, plaintiff disclosed its information to parties outside its business [F10], plaintiff's information could be ascertained by examining or reverse engineering plaintiff's product [F16], and plaintiff made disclosures of the information in a public forum [F27]. Therefore, in both cases, plaintiff's information was known in the industry or available from sources outside plaintiff's business [F105].

⇐ Plaintiff's argument emphasizing distinction F20 in *Ferranti*.

In *Ferranti*, plaintiff's information was known to competitors [F20]. This was not so in *National Rejectors*. This is a marked distinction. It shows that in *Ferranti*, defendant obtained or could have obtained its information by legitimate means [F120]. Other facts in *Ferranti* further support this: Defendant produced its information through its own independent efforts [F17]. In *National Rejectors*, by contrast, defendant may have acquired plaintiff's information through improper means [F120]: Plaintiff's former employee removed documents containing plaintiff's information [F7].

Fig. 8. CATO's arguments interpreting the significance of a distinction.

did not use (or did not need to use) improper means to obtain the plaintiff's information. That interpretation of *Ferranti* is supported also by the fact that in *Ferranti* the defendant developed its product independently, without using the plaintiff's information (factor F17 Info-Independently-Generated). In *National Rejectors*, by contrast, there was evidence that the defendants used improper means to obtain the plaintiff's information, namely, the fact that they brought product development information when they left plaintiff's employ (F7). Thus, CATO emphasizes the distinctions using its argument move: *Drawing an abstract contrast*.

Each argument is based on a different interpretation of the two cases. Both interpretations are plausible and relate to important concerns in trade secret law. Both were chosen in a strategic manner, to advance a goal to argue that cases are similar or not. The argument exchange illustrated here does not definitively resolve the question whether the given distinction matters much and whether the cases are sufficiently similar. Nonetheless, modeling context-sensitive arguments about the significance of differences among cases is a contribution to the fields of AI and Law and CBR. The given argument exchange illustrates one way in which CATO's arguments about the significance of distinctions are sensitive to the arguer's viewpoint. CATO is able to focus on different abstract interpretations of the cases being compared, depending on whether it attempts to play up or play down a distinction between them. Further aspects of the context-sensitivity of CATO's arguments are illustrated elsewhere [3,9].

Support for an abstract factor

Pro- S conclusion $P(S)$ associated with abstract factor P is supported, with respect to a reference set of base-level factors REF (typically the applicable factors in a given case or problem) iff for some pro- S factor $F \in REF$, there is a path in the Factor Hierarchy

$$F \rightarrow X_1 \rightarrow X_2 \rightarrow \dots \rightarrow X_n \rightarrow P, \quad n \geq 0$$

such that for each X_i , as well as for P , the path to it from F is strong (i.e., is made up of strong links only), or there is no strong path to it from any opposing (i.e., con- S) factor $\in REF$.

Fig. 9. Criterion for evaluating whether a conclusion associated with an abstract factor is supported.

5.3. Downplaying a distinction by drawing an abstract parallel between cases

In this section, we describe CATO's method for drawing an abstract parallel between two cases, the main way in which it can downplay the significance of a distinction. The key to this method is a heuristic criterion for selecting the *focal abstractions* used in these arguments, that is, the abstract interpretations of cases that are highlighted. In the process of selecting focal abstractions, CATO evaluates the *support* that exists for and against abstract factors in the Factor Hierarchy, using a simple way of propagating evidence along the links in the hierarchy, taking into account whether links are weak or strong. We describe this method first. In order for a conclusion associated with an abstract factor to be supported, it is not necessary that it can be proven in one sense or another. Rather, there must be evidence in favor of the conclusion. Weak evidence in favor of a conclusion can be *blocked* by strong opposing evidence. Strong evidence in favor of a conclusion however cannot be blocked. When the evidence for the two opposing conclusions associated with a given abstract factor is of equal strength, we consider both conclusions to be supported. For examples, see [1, p. 65]. A formal description of CATO's notion of support is given in Fig. 9. However, the details of how CATO evaluates support for abstract factors are unimportant. Any scheme for non-monotonic reasoning that has the properties outlined here could be used.

Now let us look in more detail at CATO's method for drawing an abstract parallel between two cases, shown in Fig. 10. Let us assume that factor D is a distinction between two given cases. CATO downplays distinction D if it can find an abstract factor P , that (a) is an ancestor in the Factor Hierarchy of D , and (b) is supported in the other case, that is, the case in which D does not apply, by one or more factors that favor the same side as D . (These factors we call *similar factors in the other case*.) Abstract factor P is used as the focal abstraction in an argument drawing an abstract parallel, provided that it is eligible to be used as a focal abstraction, meaning that it must be an Intermediate Legal Concern or a Legal Issue not linked to any Intermediate Legal Concern. This requirement was added to avoid drawing a parallel that is so broad to be unconvincing or conclusory.

To see how this criterion applies to the distinction F20 between *National Rejectors* and *Ferranti* discussed earlier, we must compare the support in each case for the abstract factors that are linked to the distinction of interest, factor F20, depicted in Fig. 11. The relevant base-level factors and abstract factors that apply in *National Rejectors* are shown on the left, those in *Ferranti*, on the right. The factors that support each relevant abstract factor, calculated according to the criterion described previously, are shown in a rounded rectangle next to the abstract factor. For example, in *National*

Drawing an abstract parallel

An abstract factor P can be used to downplay distinction D of case C_{dist} , as compared to case C_{other} , where D favors side S , by drawing an abstract parallel, if

- P is eligible for use as a focal abstraction, and
- P is an ancestor in the Factor Hierarchy of D , and
- $P(S)$ is supported in C_{other} .

Similar factors in the other case

The pro- S factors in C_{other} that are descendants of P are used as similar factors in the other case.

Eligible as focal abstraction

An abstract factor is eligible for use as a focal abstraction if it is an Intermediate Legal Concern or a Legal Issue not linked to any Intermediate Legal Concern.

Fig. 10. Criterion for selecting focal abstractions used to draw an abstract parallel.

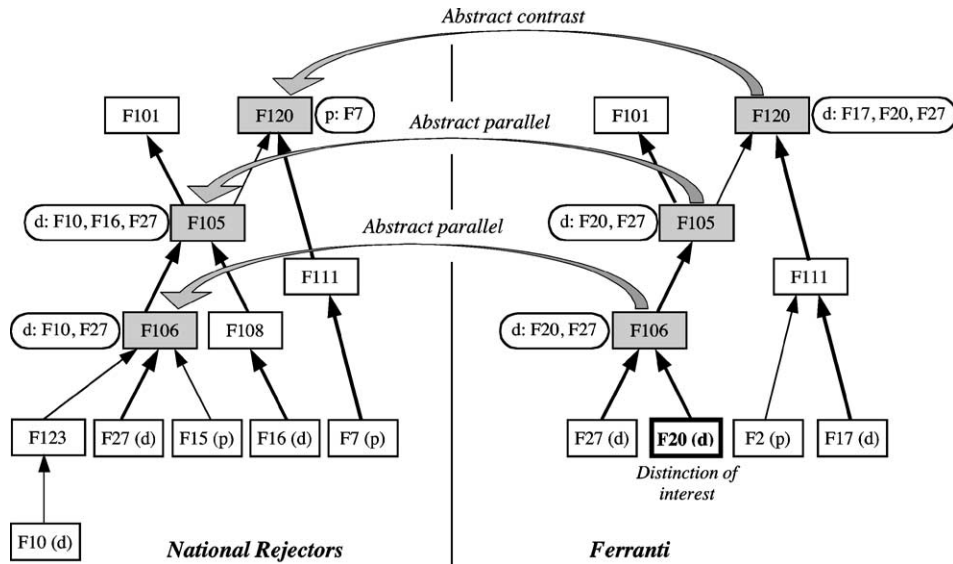


Fig. 11. Comparing the support for abstract factors in *National Rejectors* and *Ferranti* in order to select the focal abstractions for the arguments about the significance of distinction F20 in *Ferranti*.

Rejectors, the pro-defendant conclusion associated with abstract factor F106 is supported by pro-defendant factors F10 and F27. The potential support from factor F15 to this abstract factor (F106) however is blocked. The abstract factors that are eligible for use as focal abstractions, according to the criterion described above, are shown as gray boxes. Thus, we see that a parallel exists between the two cases with respect to two abstract factors, F105 Info-Known-Or-Available and the more specific F106 Info-Known. In *National Rejectors*, there is support for the pro-defendant conclusions related

to both abstract factors, as shown in Fig. 11. Both could potentially serve as focal abstractions.

When multiple abstract factors satisfy CATO's criterion for selecting focal abstractions, CATO selects those that are broad enough to maximize the set of base-level factors that can be marshaled as evidence for a contrast or parallel between the cases, but are no broader than that. By not selecting even broader abstractions, CATO avoids drawing a parallel or contrast that is so broad that it is unconvincing or that might be more easily rebutted by the opponent. In the current example, therefore, CATO prefers abstract factor F105 over the more specific F106, since F105 provides the opportunity to marshal the additional factor F16 as evidence of a parallel between the cases.

5.4. *Emphasizing a distinction by drawing an abstract contrast between cases*

In the current section, we describe how CATO emphasizes the significance of a distinction by drawing an abstract contrast between two cases. Let us assume, as before, that factor D is a distinction between two given cases. CATO uses abstract factor P (which must be an ancestor in the Factor Hierarchy of distinction D) to draw an abstract contrast, if a contrast exists between the given cases both with respect to P and with respect to abstract factors "higher up" in the Factor Hierarchy, that is, with respect to ancestors of P (see Fig. 12). The requirement that there is a contrast higher up ensures that CATO avoids drawing an abstract contrast to which an opponent could respond by drawing an even broader parallel, which would be an effective rebuttal. In order for a contrast to exist with respect to P , P must be supported in the case in which the distinction D applies by the distinction D itself and perhaps by other factors. In the other case, P must either be supported by factors that favor the opposing side, compared to D . Alternatively, it must be reasonable to argue that in the other case P is not supported by factors that favor the same side as D . For the latter condition to be met, a closed-world assumption must be warranted with respect to the given conclusion associated with P . This constraint helps avoid certain unwarranted inferences based on the absence of information. For example, even when there are no factors that indicate that the information that the plaintiff claims as its trade secret is valuable for the plaintiff's business (abstract factor F104, Info-Valuable), it makes no sense to claim the information would not be valuable. The parties would not be in court if the information had no value. By indicating that a closed-world assumption is not warranted with respect to the pro-plaintiff conclusion associated with F104, that inference can be avoided. At other times, a closed-world assumption is warranted. For example, when there is no information that indicates that the plaintiff took efforts to maintain the secrecy of the information at issue (abstract factor F102 Efforts-To-Maintain-Secrecy), arguing that none were taken is reasonable. Therefore, CATO needs to be told (by the developer of the Factor Hierarchy) for which conclusions associated with abstract factors a closed-world assumption can be made.

When an abstract factor P satisfies these criteria, a contrast between the cases can be elaborated with respect to P . In the one case, the distinction D and perhaps other factors that favor the same side (we call these additional factors *corroborating factors*) provide support for one of the conclusions associated with P . By contrast, in the other case, there is no evidence for that conclusion or even evidence for the opposite conclusion.

Drawing an abstract contrast

An abstract factor P can be used to emphasize a distinction D of case C_{dist} as compared to case C_{other} where D favors side S , if

- P is eligible as focal abstraction, and
- P is supported by D in C_{dist} and
- either $P(S)$ is not supported in C_{other} or $P(\text{opp}(S))$ is supported in C_{other} by factors that do not apply in C_{dist} and
- there is a suitable contrast between C_{dist} and C_{other} at higher levels, that is, with respect to ancestors of P in the Factor Hierarchy; specifically
 - there must be a path from P to the top of the Factor Hierarchy all of whose abstract factors (except possibly a Legal Issue at the top) are supported for S in C_{dist} , and
 - there must not be a path from P to the top of the Factor Hierarchy containing any abstract factors (except possibly a Legal Issue at the top) that are supported for S in C_{other} .
- if $P(\text{opp}(S))$ is not supported in C_{other} , then a closed-world assumption must be sanctioned with respect to $P(S)$.

Contrasting factors in the other case

All factors in C_{other} that support $P(\text{opp}(S))$ and do not apply also in C_{dist} are used as contrasting factors in the other case.

Corroborating factors in the same case

All factors in C_{dist} that support $P(S)$ (except D) and do not apply also in C_{other} are used as corroborating factors in the same case.

Fig. 12. Criterion for selecting focal abstractions used to emphasize distinctions.

Factors supporting the opposite conclusion we call *contrasting factors in the other case*.

Returning to the F20 distinction in the *National Rejectors/Ferranti* comparison, when we compare the support for abstract factors in each of the two cases, we see that abstract factor F120 Info-Legitimately-Obtained-Or-Obtainable satisfies CATO's criterion for drawing an abstract contrast (see Fig. 11). In *Ferranti*, F120 is supported by various pro-defendant factors, including the distinction F20 and corroborating factor F17. In *National Rejectors*, by contrast, F120 is supported by pro-plaintiff factor F7. CATO does not marshal pro-defendant factor F27 in *Ferranti* as a corroborating factor, even though this factor also supports F120. This factor is shared between the two cases and therefore not useful to draw a contrast.

5.5. Downplaying a distinction by showing an opposite interpretation

We have seen how CATO downplays the significance of distinctions by drawing an abstract parallel between two cases. A second way of arguing that a distinction is not important is to show that the case in which the distinction occurs has additional factors that support an interpretation of that case that is opposite to an interpretation suggested by the presence of the distinction (*showing an opposite interpretation*). This argument move is illustrated in Fig. 13, which shows CATO's arguments about the significance of a distinction in the comparison of *National Rejectors* and *Televaton*, namely, the presence in *National Rejectors* of factor F27, Disclosure-In-Public-Forum. Arguing on behalf of

National Rejectors (d)	Televation (p)
* F7 Brought-Tools (p)	* F6 Security-Measures (p)
F10 Secrets-Disclosed-Outsiders (d)	= F10 Secrets-Disclosed-Outsiders (d)
* F15 Unique-Product (p)	* F12 Outsider-Disclosures-Restricted (p)
F16 Info-Reverse-Engineerable (d)	= F15 Unique-Product (p)
* F18 Identical-Products (p)	= F16 Info-Reverse-Engineerable (d)
= F19 No-Security-Measures (d)	= F18 Identical-Products (p)
= F27 Disclosure-In-Public-Forum (d)	** F21 Knew-Info-Confidential (p)

⇒ **Plaintiff's argument downplaying distinction F27 in *National Rejectors*.**

In *National Rejectors*, plaintiff disclosed its information in a public forum [F27]. This was not so in *Televation*. However, this difference is insignificant. Even though in *National Rejectors*, plaintiff presented or published the information in a public forum [F27], defendant may have acquired plaintiff's information through improper means [F120]: In *National Rejectors*, plaintiff's former employee removed documents containing plaintiff's information [F7]. Even assuming (without conceding) that in *National Rejectors*, defendant obtained or could have obtained its information by legitimate means [F120]: This was also true in *Televation*, and yet plaintiff still won. In *Televation*, plaintiff's product information could be learned by reverse-engineering [F16], but plaintiff won nonetheless.

⇐ **Defendant's argument emphasizing distinction F27 in *National Rejectors*.**

In *National Rejectors*, plaintiff disclosed its information in a public forum [F27]. This was not so in *Televation*. This distinction is highly significant. It shows that in *National Rejectors*, plaintiff showed a lack of interest in maintaining the secrecy of its information [F102]. Other facts in *National Rejectors* also support this: Plaintiff did not take any measures to protect the information [F19]. By contrast, in *Televation*, plaintiff took efforts to maintain the secrecy of its information [F102]: Plaintiff adopted security measures [F6] and plaintiff imposed confidentiality restrictions in connection with its disclosures to outsiders [F12].

Fig. 13. CATO's arguments downplaying and emphasizing a distinction in *National Rejectors*.

the plaintiff in *National Rejectors*, CATO argues that the distinction does not amount to much, combining its two argument moves for downplaying a distinction. First, it shows an interpretation of *National Rejectors* that is opposite to that suggested by the presence of the distinguishing factor, F27. In *National Rejectors*, the fact that the plaintiff disclosed part of its product information in a public forum (F27) should not be interpreted as showing that the defendant obtained or could have obtained the plaintiff's information legitimately (F120p). That interpretation of *National Rejectors* is undercut by additional factors: In *National Rejectors*, the defendant took the plaintiff's product development documents or tools that contained the information (F7), hardly a proper way of acquiring information.

CATO combines this argument move with its other move for downplaying the significance of a distinction: it goes on to draw an abstract parallel between the cases. In both cases, there was evidence that the plaintiff's information was or could have been obtained in a proper manner, without resorting to improper means (F120d). In *National Rejectors*, the distinction F27 provides evidence for this interpretation, in *Televation*, factor F16 Info-Reverse-Engineerable. To succeed rhetorically, the argument combining these moves must take into account that they focus on opposing interpretations related to the same abstract factor. It must present them as *alternative* interpretations of *National Rejectors* that both lead to the same conclusion, namely, that the two cases are similar. The gist of the argument is that whether one argues that in *National Rejectors*, the defendant did or could have obtained the plaintiff's information legitimately (F120d), or whether one

argues that the defendant in *National Rejectors* used improper means (F120p), the cases are not different with respect to F120. To generate this argument, CATO uses a text template tailored to the situation where a single abstract factor is used simultaneously as the focal abstraction in the two different moves for downplaying a distinction. The phrase “even assuming (without conceding)” (see Fig. 13) signals that two alternative interpretations of *National Rejectors* are being presented.

CATO’s argument template was designed also to address a second rhetorical concern. In drawing an abstract parallel between the two cases on behalf of the plaintiff, CATO is forced to focus on an interpretation of the problem that is favorable to the defendant, namely, the fact that the defendant did or could have obtained the plaintiff’s information legitimately (F120d). To succeed rhetorically, the argument must stress that the plaintiff can still win, in spite of the plausibility of this (for the plaintiff) unfavorable interpretation. Therefore, CATO’s text template for drawing an abstract parallel includes phrases like “and yet plaintiff still won”. These phrases are added to the argument only when the distinguishing factor applies in the problem situation. They are not needed when it applies in the cited case, because in that situation, the abstract parallel involves a favorable interpretation of the problem rather than an unfavorable one.

All in all, CATO’s argument shown in Fig. 13 makes effective use of *Televation*. Not only does CATO downplay an apparent distinction between *National Rejectors* and *Televation*, it argues also that a weakness in the plaintiff’s position, the presence of pro-defendant factor F27, can be overcome and draws attention to a strength for plaintiff, factor F7, Brought-Tools. In its response on behalf of the defendant in *National Rejectors*, CATO elaborates a (rather striking) contrast between the cases based on a different characterization of the F27 factor (see Fig. 13), illustrating again that CATO is adept at selecting alternative characterizations of cases depending on whether its goal is to argue that the cases are similar or different.

CATO’s heuristic criterion for selecting focal abstractions to show an opposite interpretation is shown in Fig. 14. If factor D applies in a given case, then CATO uses an abstract factor P to show an opposite interpretation of D if P is linked in the Factor Hierarchy to D and is supported by applicable factors of that case that favor the opposing side, as compared to D . We call these factors *undercutting factors*. If D applies in the

Showing an opposite interpretation

An abstract factor P can be used to downplay distinction D of case C_{dist} , as compared to case C_{other} , where D favors side S , by showing an opposite interpretation, if

- P is eligible as focal abstraction, and
- P is an ancestor in the Factor Hierarchy of D , and
- $P(\text{opp}(S))$ is supported in C_{dist} ; if C_{dist} is the cited case by factors that do not apply also in C_{other} .

Undercutting factors in the same case

The con- S factors in C_{dist} that are descendants of P are used as undercutting factors in the same case. If C_{dist} is the cited case, not the problem, then only factors are used that are not shared between the cases, that is, that do not apply also in C_{other} .

Fig. 14. Criterion for selecting focal abstractions used to show an opposite interpretation.

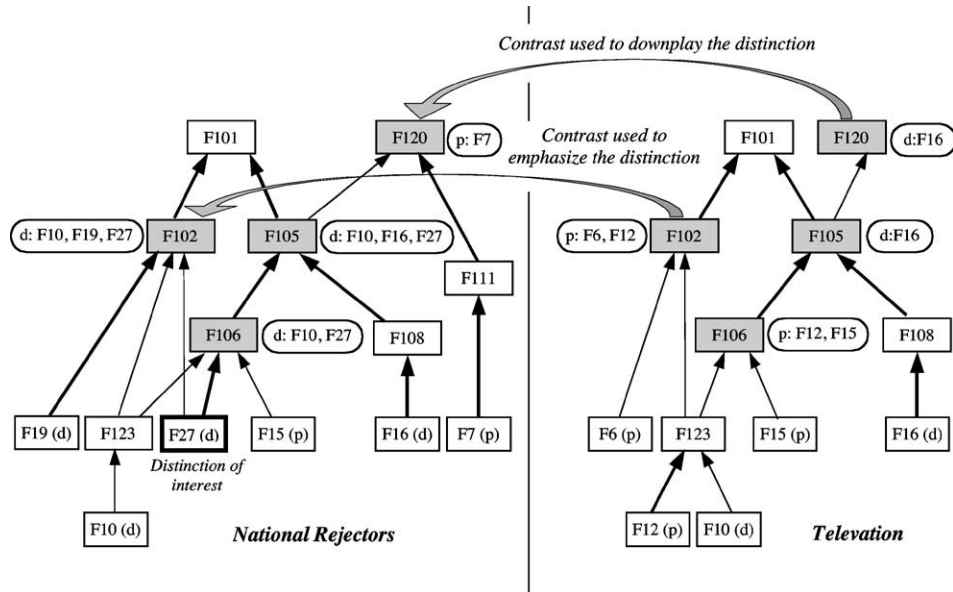


Fig. 15. Comparing *National Rejectors* and *Televation* in order to select focal abstractions that can be used in arguments about the significance of distinction F27 in *National Rejectors*.

cited case (as opposed to the problem situation), then the undercutting factors must not apply also in the problem situation, to avoid drawing attention to unfavorable factors in the problem.

To see why CATO selected F120 as the focal abstraction in its argument showing an opposite interpretation of *National Rejectors*, we must consider the support for abstract factors in *National Rejectors*, depicted on the left-hand side of Fig. 15. As before, the support for abstract factors is shown in rounded rectangles. The pro-plaintiff conclusion associated with abstract factor F120 satisfies CATO's criterion for use as an opposite interpretation. The distinction, pro-defendant factor F27, provides evidence for a pro-defendant interpretation of *National Rejectors* with respect to F120. However, pro-plaintiff factor F7, which favors the plaintiff, supports an opposite interpretation of *National Rejectors* with respect to F120. Therefore, F7 is used as an undercutting factor. The fact that the distinction F27 does not actually support the focal abstraction F120 (the potential support is blocked) is incidental; CATO's criterion neither requires nor prevents it.

Further, the pro-defendant conclusion associated with F120 satisfies CATO's criterion for being a focal abstraction in an argument drawing an abstract parallel. In *National Rejectors*, the distinction, factor F27, provides some evidence for this interpretation, in *Televation*, the interpretation is supported by pro-defendant factor F16. It may seem odd that CATO uses F120 to elaborate an abstract parallel between two cases, even though there is actually a contrast between the cases with respect to this abstract factor: As shown in Fig. 15, in *National Rejectors*, the pro-plaintiff conclusion associated with this abstract factor is supported, in *Televation*, on the other hand, the pro-defendant conclusion. However, this contrast cannot be used to distinguish the two cases, as it does not help to

show that *Televation* is stronger for the plaintiff than is *National Rejectors*. As mentioned, differences are distinctions only if they make the cited case stronger for the side that cited it than the problem.

6. Using background knowledge to select the most relevant cases

In addition to providing a computational account of important aspects of case-based legal reasoning, CATO's arguments about the significance of distinctions are the basis for novel criteria for selecting the most relevant precedent cases, an important concern in case-based legal argumentation and in CBR, generally. By and large, previous methods for selecting the most relevant cases in systems for case-based legal reasoning did not take into account the significance of distinctions. The HYPO program for example appraised users of the distinctions that cases have, but did not reason about their significance and did not take the distinctions into account when deciding what cases to present to the user as the best cases to cite in a given problem situation [7,8,13]. It makes much sense, however, to take into account the significance of differences when selecting the best cases, as we illustrate below.

In the current section, we present CATO's novel methods for assessing the relevance of past cases, in the context of a given problem situation. We illustrate how CATO's novel methods, which rely on reasoning about the significance of differences, can help to assess similarity in a more discriminating and accurate manner. In the next section, we present results from an evaluation study in which these methods were compared, in terms of their accuracy in predicting case outcomes, to an existing method used in interpretive CBR programs and to standard inductive learning techniques.

6.1. Background: A criterion for selecting the most relevant cases without using background knowledge

We first review a well-known relevance criterion that is often used when cases are represented in terms of factors, namely, the criterion used by the HYPO program [7, 8]. This criterion does not involve the use of background knowledge nor does it involve reasoning about the significance of distinctions. We then discuss how CATO's techniques for reasoning about the significance of distinctions can best be integrated with this method.

HYPO's main criterion for selecting the most relevant cases was designed with the following constraint in mind [8, Ch. 9]: An interpretive CBR program must select the most relevant cases according to criteria that a human expert would recognize as plausible and that provide an explanation or argument that the selected cases are more relevant than those that were not selected. This constraint implies at minimum that the program's decisions should not rely on arbitrary factor weights or ways of calculating factor weights that cannot be justified in an argument [13]. They should also not rely on simple counts of the similarities and differences. HYPO's criterion for selecting the most relevant cases, called *best untrumped cases* or *BUCs*, observes this criterion. As illustrated below, CATO's relevance criteria also adhere to this constraint. The *BUCs* are cases that:

- satisfy a minimum criterion for being *citable* in an argument, namely, that they share with the problem at least one factor that favors the side that won the case, and
- of all available cases, have a maximum set of factors shared with the problem (these cases are *most on point*, in that no other case won by the same side shares a more inclusive set of factors with the problem), and
- are *untrumped*, meaning that there is no trumping counterexample, a case won by the opponent that is *more on point* in that it shares a more inclusive set of factors with the problem.

6.2. Two methods for applying background knowledge to select the most relevant cases

In this section, we discuss how CATO's methods for reasoning about the significance of differences can best be combined with (components of) the *BUC* criterion, which focuses on the similarities that cases share with a given problem. This integration can be achieved in different ways, depending on how much weight one gives to the absence of significant distinctions in one case, relative to the presence of extra similarities in another. *A priori*, it is not clear whether and how the two constraints should be traded off against each other. We illustrate the pros and cons of two different ways of doing so.

To take into account the significance of distinctions, one could simply add a further constraint to the *BUC* criterion that the cases must not have significant distinctions. Using CATO's background knowledge in this manner sometimes leads to more discriminating relevance assessment, as the following example illustrates. In *National Rejectors*, the problem situation presented above, the court held for the defendant, relying primarily on *Yokana*, a case won by the defendant: "[W]e do find some significant parallels between the facts of this case and those of *Midland-Ross Corporation v. Yokana* (D.C.N.J.), 185 F.Supp. 594." The court's decision for the defendant and the central role that *Yokana* played in its reasoning cannot be explained solely on the basis of the *BUC* criterion. In *National Rejectors*, the set of best untrumped cases, that is, cases in CATO's Case Database that satisfy the *BUC* criterion, includes 4 cases won by the plaintiff and 10 cases won by the defendant, including *Yokana*. Thus, based on the *BUC* criterion, we conclude that both sides have a strong argument. The fact that more pro-defendant cases than pro-plaintiff cases satisfy the *BUC* criterion makes no difference, since simple counts of relevant cases cannot be used as a premise in a convincing legal argument.

If we employ CATO's methods for reasoning about the significance of the distinctions, however, we can explain why the court in *National Rejectors* focused on *Yokana* and held for the defendant. All 14 cases that satisfy the *BUC* criterion have (multiple) distinctions. A number of the pro-defendant cases, including *Yokana*, do not have *significant* distinctions, as can be ascertained using CATO's methods: CATO finds arguments downplaying the significance of all of *Yokana*'s distinctions and no arguments emphasizing them. On the other hand, all four pro-plaintiff cases that satisfy the *BUC* criterion have at least one significant distinction, according to CATO's criterion. Thus, the example illustrates that one way of using CATO's arguments about the significance of distinctions can lead to more discriminating relevance assessment, namely, by narrowing down the set of case that satisfy the *BUC* criterion to include only the cases that have no significant distinctions. We will call this method *BUC/NoSignDist*.

<i>Goodrich</i>	<i>Motorola</i> (d)
= F2 Bribe-Employee (p)	= F2 Bribe-Employee (p)
= F4 Agreed-Not-To-Disclose (p)	= F4 Agreed-Not-To-Disclose (p)
= F5 Agreement-Not-Specific (d)	= F5 Agreement-Not-Specific (d)
= F6 Security-Measures (p)	= F6 Security-Measures (p)
	** F20 Info-Known-To-Competitors (d)
	** F27 Disclosure-In-Public-Forum (d)
<i>Goodrich</i>	<i>Eastern Marble</i> (p)
F2 Bribe-Employee (p)	= F4 Agreed-Not-To-Disclose (p)
= F4 Agreed-Not-To-Disclose (p)	= F5 Agreement-Not-Specific (d)
= F5 Agreement-Not-Specific (d)	= F6 Security-Measures (p)
= F6 Security-Measures (p)	* F15 Unique-Product (p)

Fig. 16. Comparison of the *Goodrich* problem situation to two precedent cases, *Motorola* (top) and *Eastern Marble* (bottom). “*” means an insignificant distinction, “**” indicates a significant distinction.

A second example illustrates a potential limitation of this method and presents an alternative method. Implicit in the *BUC* criterion (and by extension, the *BUC/NoSignDist* criterion) is an assumption that the cases that are most similar to a given problem have to be found among the cases that have a maximum set of similarities. However, there may be times when the cases with the most inclusive set of similarities have significant distinctions, whereas some cases with fewer similarities do not. On balance, those cases may be more attractive, but they will not satisfy the *BUC* criterion.

Compare *Goodrich*, a case won by the plaintiff, to two precedent cases: *Motorola*, won by the defendant, and *Eastern Marble* won by the plaintiff. Comparisons of the factors of each of the precedent cases to those of *Goodrich* are shown in Fig. 16. *Motorola* is *more on point* than *Eastern Marble*: it has all factors that *Eastern Marble* shares with *Goodrich* (namely, F4, F5, and F6) and shares an additional factor with *Goodrich*, F2, Bribe-Employee. If the plaintiff in *Goodrich* drew an analogy between *Goodrich* and *Eastern Marble*, the defendant could “trump” plaintiff’s point by citing *Motorola* as a *more on point counterexample* [7,8]. Consequently, *Eastern Marble* does not satisfy the *BUC* criterion. As it turns out, *Motorola* is the only case in CATO’s Case Database that satisfies this criterion. Therefore, the use of the *BUC* criterion would lead one to predict, incorrectly, that *Goodrich* would be won by the defendant. The *BUC/NoSignDist* criterion introduced above also does not fully explain the pro-plaintiff result in *Goodrich*. No case satisfies this criterion. In particular, *Motorola*, the only case that satisfies the *BUC* criterion, has two significant distinctions and thus does not qualify.

On reflection, it is not appropriate to consider *Motorola* to be more relevant than *Eastern Marble*. *Motorola* has two distinctions, compared to *Goodrich*, both of which are significant: CATO finds arguments emphasizing them but finds no arguments to downplay them. By contrast, *Eastern Marble*’s sole distinction, F15, Unique-Product, is insignificant. But if the absence of factor F2 in *Eastern Marble* does not amount to a significant distinction (in fact, it does not amount to a distinction at all, as shown in Fig. 16), then why would the presence of that factor in *Motorola* be reason to consider *Motorola* to be more relevant than *Eastern Marble*, as the *BUC* method would indicate? It is desirable and appropriate to attribute less weight to the presence of the extra factor F2 in *Motorola* and more weight to the fact that it has significant distinctions, whereas *Eastern Marble* does not.

We introduce a new criterion for selecting the best cases that considers the absence of significant distinctions in one case to be more important than the presence of extra similarities in another. In applying this criterion, one first selects the cases without significant distinctions. Then from among those, one selects the ones that are most on point. We call this method *NoSignDist/BUC*. Under this criterion, *Eastern Marble*, not *Motorola*, is the more relevant precedent in *Goodrich*. Therefore, on the basis of this new criterion we can explain the pro-plaintiff outcome in *Goodrich*.

In sum, the two examples illustrate two different ways in which CATO's method for making context-sensitive arguments about the significance of distinctions might be used to improve the assessment of the relevance of cases. In the next section, we present results of an evaluation study investigating to what extent these examples are representative of a wider range of cases.

7. Evaluation study 1: Predictive accuracy and the use of background knowledge

One way to evaluate CATO's arguments, or its judgments of case relevance, would be to gather a representative sample of CATO's output, and compare it against arguments or relevance judgments produced by legal experts or competent novices. An evaluation of this type is presented in the next section. The downside of this type of evaluation is that it is very labor-intensive. One needs to have experts and/or competent novices analyze a problem, read the opinions or squibs of relevant cases, and write a legal brief. Even in its simplest form, this task takes several hours. Ideally, one would have multiple graders assess the results. It is often infeasible to use this method to do comparative evaluations of different ways of generating arguments.

A useful supplementary approach is to look at how well a program predicts the outcome of cases, based on its arguments or judgments of case relevance. Good predictive performance would inspire confidence that the arguments made by the program are good arguments that have some relation to the reality of legal reasoning. Perhaps a focus on predicting case outcomes is surprising, given that the primary goal of interpretive CBR programs like CATO is to make arguments, not predictions. However, argumentation and prediction are closely tied in programs for case-based argumentation such as CATO. Typically, in the process of constructing arguments, these programs select the most relevant cases that each side can cite to support the strongest possible argument. To apply them to the task of predicting case outcomes, all that needs to be done is equip them with a simple decision rule, namely, that a given side wins if the most relevant cases all favor that side. Given this close tie between argumentation and prediction, it would reflect positively on an argumentation model if it could be applied to make accurate predictions of case outcomes. On the other hand, not every procedure that leads to good classification performance necessarily produces good arguments. We therefore recommend using this evaluation procedure in combination with methods that focus on the rhetorical structure and quality of arguments, as we do in the current paper.

We conducted an experiment to investigate whether CATO's background knowledge enables it better to predict the outcome of cases, as compared to methods for assessing the relevance of cases that do not involve the use of background knowledge, such as the

BUC criterion [7,8]. What level of performance to expect is an open question. Predicting the outcome of legal cases is a notoriously difficult problem. In CATO, this difficulty is compounded by the fact that the case database contains cases from different jurisdictions and time periods and involving a wide range of procedural settings. Further, some cases were decided based on the Restatement 1st of Torts, others under the Uniform Trade Secrets Act. With such a heterogeneous mix of cases, one does not expect to see perfect classification performance. Therefore, in the evaluation study, we compared CATO's methods not only against interpretive CBR methods not relying on background knowledge, but also against standard inductive learning algorithms. These programs do not generate arguments, so in that sense they are not comparable, but they do help to get a sense of how easy or difficult it is to predict the outcome of the cases in CATO's database.

7.1. Evaluated methods

The experiment had two main goals: to find out how CATO's methods compare to (a) the *BUC* criterion [7,8] described above, a well-established relevance criterion for selecting the best cases to cite when working with factor-based representations of cases which does not involve the use of background knowledge and (b) three standard inductive learning algorithms. In addition, the experiment had two more specific goals. First, we wanted to know how CATO's methods can best be employed to predict case outcomes—in particular, how much weight to give to the absence of significant distinctions in one case, compared to the presence of extra similarities in another, an issue that was illustrated in the two examples shown above. Second, we wanted to investigate whether any improvement seen due to the use of background knowledge could also have been obtained by simpler methods that take into account whether cases have distinctions but do not reason about the significance of the distinctions.

We evaluated the performance of 7 different CBR methods for predicting case outcomes, each based on a different criterion for selecting the most relevant cases. Each method uses its relevance criterion, *R*, as follows:

```

find all cases that satisfy relevance criterion R
if there are cases that satisfy R and all of them were won by the same side
then predict that this side wins
otherwise, abstain

```

The relevance criteria, shown in Table 1, used can be divided into three groups. The first “group” comprises a single criterion, the *BUC* criterion discussed above, which serves as a control against which to evaluate the improvement (if any) due to CATO's background knowledge. This method considers only the similarities of cases. It does not take into account the significance of distinctions or whether the cases have distinctions.

The second group contains three methods that rely on CATO's background knowledge and reasoning about the significance of distinctions. They combine reasoning about the significance of distinctions with (components of) the *BUC* criterion. These methods differ in the weight that they give to extra similarities in one case, as compared to the absence of significant distinctions in another. First, the *NoSignDist/Cit* criterion considers

Table 1
Relevance criteria used for selecting the best cases

A. Criterion used as control		
1.	<i>BUC</i>	Best untrumped cases
B. Criteria that require that cases have no significant distinctions		
2.	<i>NoSignDist/Cit</i>	Citable cases that have no significant distinctions
3.	<i>NoSignDist/BUC</i>	Of the citable cases without significant distinctions, those that are most on point and untrumped, by citable cases without significant distinctions
4.	<i>BUC/NoSignDist</i>	Of the best untrumped cases, those that have no significant distinctions
C. Criteria that require that cases have no significant distinctions—these criteria do not involve the use of background knowledge		
5.	<i>NoDist/Cit</i>	Citable cases that have no distinctions (mirrors criterion 2)
6.	<i>NoDist/BUC</i>	Of the citable cases without distinctions, those that are most on point and untrumped, by citable cases without distinctions (mirrors criterion 3)
7.	<i>BUC/NoDist</i>	Of the best untrumped cases, those that have no distinctions (mirrors criterion 4)

as most relevant those cases that (1) have no significant distinctions, meaning that any distinctions that they may have can be downplayed and cannot be emphasized and (2) satisfy HYPO's minimum criterion for a case being *citable* in an argument [7,8], namely, that they share with the problem at least one factor that favors the side that won that case. The *NoSignDist/Cit* does not place much emphasis on similarities, requiring only that cases have a minimal set of similarities in common with the problem. The second criterion, *NoSignDist/BUC*, which was illustrated in the *Goodrich* example, places more emphasis on similarities than the previous one, but still considers the absence of significant distinctions to be of greater importance than the presence of extra similarities. The cases that satisfy the *NoSignDist/BUC* criterion can be selected in two steps: first, one selects the cases that have no significant distinctions, as defined above. From among those cases, one selects the ones that satisfy the *BUC* criterion. The third criterion, *BUC/NoSignDist*, finally, gives more weight to the presence of extra similarities than the previous two. Under this criterion, cases are required to have maximum sets of similarities *and* not to have significant distinctions. This method was illustrated in the *National Rejectors/Yokana* example. When using the *BUC/NoSignDist*, the steps for selecting the best cases are reversed, as compared to *NoSignDist/BUC*: One first selects the cases that satisfy the *BUC* criterion and then from among those, selects the ones that have no significant distinctions. Both methods, *NoSignDist/BUC* and *BUC/NoSignDist*, are justifiable *a priori*, as discussed in the context of the two examples given above.

The third group contains methods that take into account whether or not a case has distinctions, but do not rely on reasoning about the significance of the distinctions. We include these methods to ascertain that any improvements that we may see, due to CATO's arguments about the significance of distinctions, could not have been obtained in a more straightforward manner. Thus, for each method that requires that cases have *no*

significant distinctions, we include a corresponding method that requires that the cases have *no* distinctions. For example, the *NoDist/Cit* criterion focuses on cases that have no distinctions and satisfy the minimum criterion for being citable in an argument. It is analogous to the *NoSignDist/Cit* criterion.

In addition to the 7 CBR methods, we evaluated the performance of three standard inductive learning methods, ID3, k-NN, and Naïve Bayes, described in [48], in order to get a sense of how difficult it is to classify the cases in CATO's database.

7.2. Experimental procedure

The experiment was carried out with a case database that contained 184 trade secrets cases. Of these cases, 107 (58%) were won by the plaintiff, 77 (42%) by the defendant. The cases are represented using a set of 26 factors, half of which favor the plaintiff, half of which favor the defendant. On average, each case has 4.1 ± 1.6 applicable factors. Each factor applies in 29 ± 22 cases. Leave-one-out cross-validation was performed for each of the 10 methods described above. For *k*-NN, we used $k = 5$ and computed the similarity of two cases as the number of shared factors minus the number of factors that applied in one case but not in the other. For each method we computed the percentage of cases that were classified correctly, incorrectly, and were not classified. The results are shown in Fig. 17. As a baseline to interpret the performance of the classification methods, one might consider that a method that predicts for every case that the plaintiff wins will be 58% accurate.

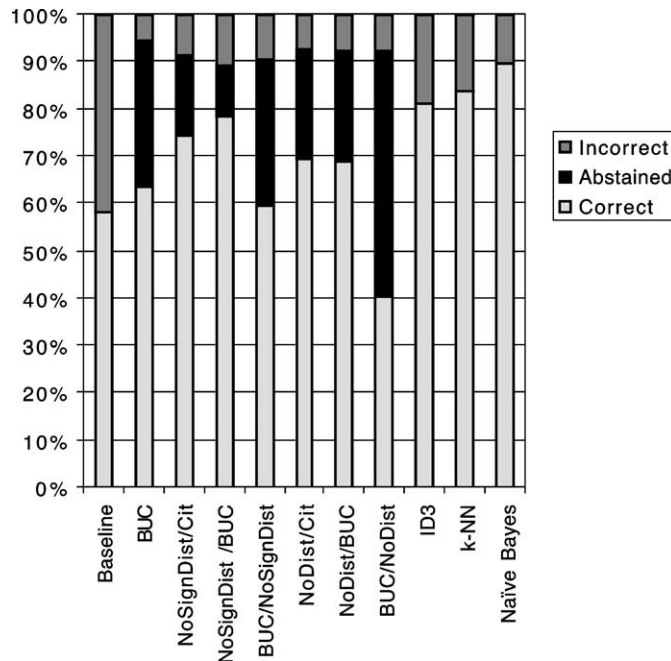


Fig. 17. Classification performance of 10 CBR and machine learning methods.

7.3. Results

We found substantial performance differences between the seven CBR methods (see Fig. 17). The differences were due mostly to the frequency with which the methods make predictions and to a lesser degree to the accuracy of the predictions. The percentage of cases for which the methods generate predictions ranges from 48 to 89%, or 69 to 89% if we disregard one outlier. The accuracy of the predictions ranges from 84 to 92%.

We used McNemar's test to determine whether the differences that were observed between the methods are statistically significant, as was recommended in a number of recent papers [31,60]. This test is essentially a Sign Test and compares the number of cases on which one algorithm does better against the number of cases on which the other algorithm does better. We consider a correct prediction to be better than no prediction, which in turn we consider to be better than an incorrect prediction. Using McNemar's test we ascertained that four of the six CBR methods performed significantly better than the baseline. The exceptions were *BUC/NoSignDist*, which was only marginally better, (#BA = 71, #AB = 49, $p < 0.1$)⁵ and *BUC/NoDist*, which performed on par with the baseline.

The *BUC* criterion, the standard against which to measure improvement, classified 69% of the cases, abstaining on the rest. When it made a prediction, it was 92% accurate. With 5% incorrect predictions it had the lowest error rate of all methods tested, including the inductive learning algorithms. Of all CBR methods tested, *NoSignDist/BUC* performed best. It classified 89% of the cases and abstained on 11% of the cases. The predictions that it made were 88% accurate. The difference between this method and the *BUC* criterion used as control was statistically significant (#BA = 17, #AB = 33, $p < 0.05$). A second method that applied background knowledge, *NoSignDist/Cit*, also performed better than *BUC*, but the difference was not statistically significant (#BA = 21, #AB = 34, $p < 0.11$). The only method using background knowledge that did worse than *BUC* is *BUC/NoSignDist*. The difference with *BUC* is marginally significant (#BA = 38, #AB = 24, $p < 0.1$).

The CBR methods that take into account the significance of distinctions did better than the corresponding methods that take into account the presence or absence of distinctions without reasoning about their significance. The difference between *BUC/NoSignDist* and its corresponding method, *BUC/NoDist*, was statistically significant (#BA = 7, #AB = 40, $p < 0.0001$), but, as mentioned, neither of these methods was clearly better than the baseline. The difference between *NoSignDist/BUC* and *NoDist/BUC* was marginally significant (#BA = 10, #AB = 21, $p < 0.1$). The difference between *NoSignDist/Cit* and *NoDist/Cit* was not statistically significant.

Of the inductive learning programs, Naïve Bayes performed best, classifying 90% of the problems correctly and the rest incorrectly. *k*-NN was 84% accurate, ID3 81%. The difference between Naïve Bayes and the best-performing CBR method, *NoSignDist/BUC*, was statistically significant (#BA = 27, #AB = 10, $p < 0.01$). On the other hand, the difference between the *NoSignDist/BUC* and the other two inductive learning programs was not statistically significant. These two inductive learning programs make more correct predictions than *NoSignDist/BUC*, but also commit more errors.

⁵ #BA indicates the number of cases on which one method did better than the other. #AB indicates the number of cases in which it was the other way around.

7.4. Discussion

On the basis of these results, the question whether CATO's background knowledge helps to make better predictions of the outcome of trade secrets cases, compared to CBR methods that do not apply background knowledge, can be answered affirmatively. *NoSignDist/BUC*, of the methods that employ background knowledge the one that performed best, was significantly better than the *BUC* method used as control, which does not apply background knowledge.

Not all CBR methods that applied background knowledge were equally effective. The methods that give more weight to the absence of significant distinctions in one case, as compared to the presence of extra similarities in another (*NoSignDist/BUC* and *NoSignDist/Cit*) did better than the methods that give more weight to extra similarities (*BUC* and *BUC/NoSignDist*). As mentioned, the difference between *NoSignDist/BUC* and *BUC* is statistically significant. Also, both *NoSignDist/BUC* and *NoSignDist/Cit* did significantly better than *BUC/NoSignDist* (#BA = 4, #AB = 35, $p < 0.0001$ and #BA = 10, #AB = 38, $p < 0.0001$). These results lend support to the argument made in context of the *Goodrich* example, namely, that an extra similarity of one case that is not a significant distinction of another case should not carry too much weight in deciding which of those cases is more relevant. The results indicate also that the *Goodrich/Motorola/Eastern Marble* example is more typical than the *National Rejectors/Yokana* example. What is unusual about the *Goodrich* example however is the fact that the *BUC* criterion (the method used as control) generates an incorrect prediction. As mentioned, the *BUC* criterion had the lowest error rate of all methods tested. Far more frequently, when *NoSignDist/BUC* (the best-performing CBR method) did better than *BUC*, it was because *BUC* abstained, having found cases won by both sides that satisfied the criterion. Thus, the main reason that CATO's background knowledge enables it to perform better on the prediction task is not because it helps to avoid errors but because it affords more discriminating assessment of the similarity of cases.

The two CBR methods that give more weight to the presence of extra similarities, as compared to the absence of significant differences, abstained on a high number of cases, 31% for both methods. However, the reason for abstaining was different. When *BUC* abstained, 98% of the time it was because there were cases won by both sides that satisfied the criterion. On the other hand, when *BUC/NoSignDist* abstained, 88% of the time it was because no cases satisfied the criterion. While *BUC/NoSignDist* may seem attractive from a theoretical viewpoint, as illustrated in the *National Rejectors/Yokana* example above, as a practical matter it puts the bar too high. As it tries to have the best of both worlds, namely, cases that have a maximum set of the factors that apply in the problem *and* that do not have significant distinctions, it frequently comes up empty, as illustrated in the *Goodrich* example.

The results of the experiment suggest, but do not prove conclusively, that the improvement in classification performance due to the use of background knowledge could not have been achieved with simpler CBR methods that consider only whether there are distinctions but do not reason about the significance of the distinctions. The main evidence is the fact that the difference between *NoSignDist/BUC*, the best-performing method that reasons about the significance of distinctions, and *NoDist/BUC*, the corresponding method

that considers only whether there are distinctions, was marginally significant. It will be interesting to investigate whether the advantage of reasoning about the significance of distinctions will be more pronounced with smaller sets of cases. We predict they will be, as the requirement that cases have no distinctions is likely to be too strict when fewer cases are available.

As expected, the results of the inductive learning programs confirm that the cases in CATO's database present a challenging classification problem. The greatest surprise was the good performance of Naïve Bayes. Apparently, this method is better able than the CBR methods to take advantage of the information available in the cases. The other inductive learning programs did not perform significantly better than the best CBR method. Perhaps they would have done better than they did if a significant effort had been made to tune them, but this possibility was not pursued. (Naïve Bayes however does not have any parameters that can be tuned.) The better performance of Naïve Bayes may be due to the fact that the CBR methods are subject to more stringent requirements: they must not only predict the outcome of a given problem, but must also explain its predictions by means of persuasive legal arguments. The inductive learning programs are not subject to this requirement. The fact that Naïve Bayes outperformed the CBR methods indicates that there is room for improvement. Perhaps these methods miss opportunities for emphasizing or downplaying the significance of certain distinctions. Perhaps what is needed is a more context-sensitive way of trading off the absence of significant differences and the presence of extra similarities—the methods tested in the experiment rely on static ways of doing so. If more context-sensitive methods can be found, it will be interesting to compare their performance to that of Naïve Bayes. As it stands, if the sole objective were to achieve high classification accuracy, Naïve Bayes would be the method of choice. However, our primary interest is argumentation and in what the experiment says about CATO's argument about the significance of distinctions. The fact that CATO's arguments add predictive value, or help explain the outcome of cases, as compared to CBR methods that do not apply background knowledge, inspires confidence that they are sound. Apparently, at an aggregate level, arguments about the significance of distinctions capture something significant about the way legal cases are decided.

8. The CATO instructional environment

Let us now turn to the question whether CATO's argumentation model can be used to support effective computer-based instruction. CATO is an intelligent learning environment, designed to help beginning law students learn basic skills of making arguments with cases. CATO supports practice of two main tasks, theory-testing (induction) and written argumentation (reasoning by analogy). In the theory-testing task, the students are given a general statement about what the law might say and then test this statement against past cases, which they retrieve from CATO's Case Database. In the argumentation task, students create a written argument about a problem situation, supported by cites to past cases selected from CATO's database. In the process of creating their argument, they study argumentation examples generated dynamically by CATO. The

CATO environment helps to make the students' tasks more manageable: it suppresses some of the distracting complexity that would complicate the use in the same instructional context of standard legal research tools and resources. The CATO environment uses its argumentation model primarily to generate examples and to make explicit the underlying structure of arguments. Unlike many intelligent tutoring systems [5,71], CATO does not use its model as a standard against which to evaluate students' solutions, with the exception of two small subtasks. If successful, CATO will address an urgent need in American law schools, since legal writing instruction is important and requires a considerable amount of (human) resources, often more than law schools are able to provide.

We evaluated how effective instruction with CATO is, compared to the best traditional way of teaching argumentation skills. In the United States, argumentation skills are taught in law schools classrooms by engaging students in a Socratic dialogue, in which a professor leads students through interesting arguments, case comparisons, and hypothetical problems to explore the ramifications of landmark cases and other legal sources.

The experiment evaluating CATO is interesting from the perspective of research in the field of AI and Education, because it tests the effectiveness of an example-based, component-wise approach to teaching argumentation skills, a relatively low-cost way of developing an instructional system for a difficult domain. The experiment is also interesting from a more theoretical perspective: Example-studying can be an effective learning method [28]. A number of example-based intelligent learning environments have been developed [25,29,34,63], but to our knowledge, none for teaching argumentation skills. While a number of other instructional systems have been developed for the legal domain [26,49,58,64,66], they have focused only to a limited extent on argumentation and have not focused on reasoning with cases. Further, those that provide feedback to students do so on the assumption that the correctness of a solution offered by the student can be determined by objective criteria, a strong assumption to make in many parts of the legal domain. In this respect these programs are rather different from CATO, which does not make this assumption, using its model to provide examples (for the most part) rather than feedback. Many intelligent learning environments for domains other than the law use case-based reasoning for a large variety of purposes, such as teaching by story telling [32,33,38,61], student modeling [70], retrieval by students of software code for re-use [27], semi-automatic grading of student work [68], initiating discussion among experts [37], and generating problems of appropriate difficulty [72]. Yet none of these systems engaged students in a process of comparing and contrasting cases in order to make arguments about how a problem should be decided, as does CATO.

The study evaluating the effectiveness of the instruction with CATO, apart from addressing an important question of interest to the research field of AI and Education, also involves further evaluation of the CATO model. First, if CATO's arguments turn out to be pedagogically useful, that result would provide further evidence that they capture important aspects of the arguments that human experts make. Second, during the evaluation study we compared CATO's arguments in a more direct way to arguments made by competent novices, by having a legal writing instructor grade written argumentation assignments completed by each.

CATO Database	— Contains textual summaries (squibs) and factor sets for 147 trade secrets cases. The query language lets students retrieve cases with any boolean combination of factors.
Factor Browser	— Provides information about CATO's set of 26 factors for trade secrets law.
Case Analyzer	— Lets students compile a list of applicable factors for a case and generates feedback, comparing student's list of factors to that stored in the CATO Database for same case.
Argument Maker	— Presents argumentation examples in the context of the student's on-going work. Also, conducts mini-dialogs to give students practice in identifying distinctions.
Issue-Based Argument Window	— Presents examples of arguments (organized by issues) with multiple cases selected by students.
Squib Reader	— Displays squibs of retrieved cases.

Fig. 18. Tools offered by the CATO program.

In the current section, we present an overview of the CATO intelligent learning environment and describe in some detail the theory-testing and argumentation tasks that it supports. In the next section, we present the results of the evaluation study.

8.1. Overview of the instruction with CATO

The CATO learning environment comprises a set of six tools, listed in Fig. 18, including a database of cases, tools that display information about retrieved cases that helps in interpreting the cases, and tools for displaying argumentation examples. At the time of the evaluation study of CATO, the Case Database contained 147 trade secrets cases. For each case, the database contains a list of applicable factors and a “squib”, that is, a short textual summary of the court's opinion, the official record in which the court describes the case facts and explains its decision. The fact that students have access to factor representations of cases and squibs is an important difference with standard legal information retrieval systems, which contain only the full opinions of cases. Another key difference is that students can ask CATO to generate argumentation examples. Upon students' request, CATO's Argument Maker, shown in Fig. 19 together with the Case Database window, displays examples of five different Basic Argument Moves, annotated to reveal the structure of the argument. Another tool, the Issue-Based Argument Window, presents, also upon students' request, arguments about a problem with any set of cases retrieved from the Case Database. Finally, CATO provides feedback to students related to two subtasks of the argumentation task: CATO's Case Analyzer lets students analyze a problem situation by selecting applicable factors and provides feedback on their choice of factors. Also, CATO's Argument Maker conducts a *Mini Dialog* to help students learn to distinguish cases, in particular, to help them to learn to differentiate between mere differences and real distinctions. Examples are shown in [1,2]. Recent work has shown that students' learning improves when role playing and courtroom drama are added to the Mini Dialogs [11].

We organized the instruction with CATO, 7.5 hours total, around a rather traditional casebook chapter for trade secrets law—many law school textbooks are casebooks. The Casebook contains a short introduction to trade secrets law, the opinions of five leading trade secrets cases, and a number of case squibs. The Casebook also contains a

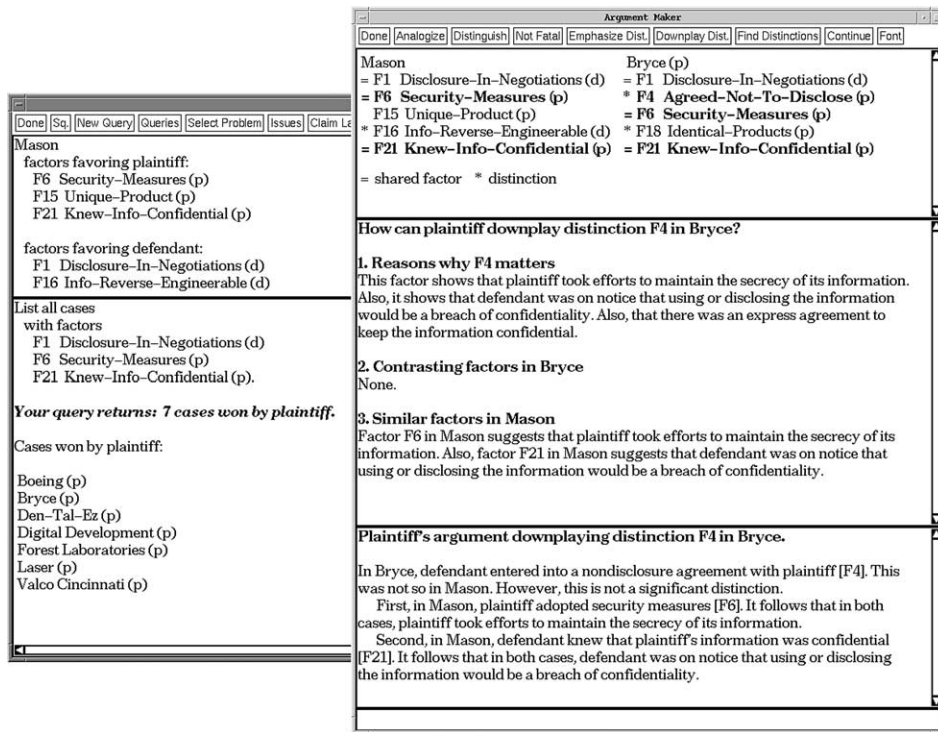


Fig. 19. CATO's Argument Maker presents dynamically-generated argumentation examples.

number of argumentation problems designed to be carried out with CATO or without.⁶ We also created a set of a set of four Workbooks, which contain brief descriptions of the argumentation skills addressed by the CATO curriculum, tutorials on how to use CATO's tools, and instructions on how to use CATO to carry out the theory-testing and argumentation tasks posed in the Casebook.

8.2. First task students practice with CATO: Theory testing

In theory-testing problems, students test a general hypothesis of what the law might say against the cases in CATO's Case Database. They revise the theory if it turns out not to be consistent with the cases. The Workbook instructions outline a 4-step process. First, the students are presented with a statement describing a general class of fact scenarios. The two scenarios that were tested during the instruction with CATO are shown in Fig. 20. They are asked to predict which side would win in the given kind of fact scenario, the plaintiff or the defendant. Second, the students frame a query to retrieve cases from CATO's

⁶ The argumentation problems had been developed using a module of CATO specifically designed to this purpose. For more information on experiments to select small sets of cases for use in argumentation instruction semi-automatically, see [10].

Suppose a defendant to whom confidential information was disclosed knew that the information was confidential, but there was not written nondisclosure agreement. Is the defendant under an obligation not to use or disclose the information?

Suppose a plaintiff disclosed confidential information in negotiations with a defendant, but the plaintiff never told the defendant that the information was confidential and there was no written nondisclosure agreement. Is the defendant under an obligation not to use or disclose the information?

Fig. 20. Theories students explored using CATO.

Case Database that are relevant to their theory. This means that they must “translate” the theory into the language of factors. This translation step is rather direct, since the theory is essentially stated at the level of factors. Third, the students consider whether the retrieved cases confirm or disconfirm the theory. To this end they determine whether any retrieved cases are inconsistent with their prediction—easy to do since CATO sorts the retrieved cases based on outcome, whether the plaintiff or the defendant won. If there are seemingly inconsistent cases, the students must investigate whether these cases invalidate the theory or whether they are outside the realm of the theory and hence pose no threat. In determining what a case says about a theory, it is useful to construct an explanation for the case’s outcome. As a first step, the students can look at the list of applicable factors of a case, displayed by CATO upon request, to get an initial idea of what the case means. They can evaluate whether the case has any additional factors not mentioned in the query that might be reason to consider the case to be outside of the realm of the theory. They can read the squib of the case in order to get more detailed information and to confirm the initial interpretation of the case. If there are inconsistent cases that cannot be explained away, students modify the theory as needed, for example by adding conditions, and they can then repeat the steps to test the modified theory.

The theory-testing exercises that students carry out with CATO are a useful type of exercises. At first blush, a focus on formulating and testing generalizations may seem odd in the context of teaching a process of reasoning by analogy, especially since the need for analogical reasoning in the legal domain is motivated often by saying that there are few useful generalizations in this domain and that generalizations often have exceptions. In part, the theory-testing exercises are meant to illustrate exactly that point, that is, that it is difficult to form generalizations that are perfectly consistent with all cases. But there is no question that the generalizations themselves are important. Even though cases are the superior authorities in common law domains, common law rules do exist and are used by legal professionals [24], for example, to organize arguments. Thus, attorneys often formulate theories and test them against available cases.

As compared to working with real-world resources for legal research and analysis, this type of exercise is simplified in CATO, but simplified in a pedagogically useful way. CATO helps make the task more manageable than if it were done with a standard legal information retrieval system. Available systems offer a wide range of very extensive case law databases, which contain case opinions but contain neither factor representations of cases nor squibs, as does CATO’s Case Database. These databases can be searched using full-text retrieval

methods.⁷ While they are extremely useful tools, it can be far more difficult to find cases that are relevant to a given theory with a standard legal information retrieval system than it is with CATO, at least for those areas of the law for which a CATO Case Database exists. For example, to find cases that have factor F1 Disclosure-In-Negotiations, one might try a query like “disclosure (disclosed) in negotiations”. However, if a case matched this query, there is a likelihood but by no means a guarantee that the corresponding factor actually applies. The terms mentioned in the query may occur in other contexts than anticipated. Further, even when the given factor does apply in a given case, the terms mentioned in the query to describe the factor may not occur in the case opinion, for various reasons. These problems occur frequently and are compounded when—as happens frequently when using CATO—one runs queries that involve multiple factors. Thus, one must read and interpret the opinions of the retrieved cases in order to find out which factors apply and whether they were won by the plaintiff or the defendant. This process can be very time-consuming, since the case opinions tend to be long documents and are not easy to read for novices. Quite often, the net result of that effort will be that the given case does *not* have the factors that one is looking for and therefore is *not* relevant to the problem at hand!

In the CATO environment, on the other hand, much of this distracting complexity is removed. Retrieved cases are guaranteed to have the factors specified in the query and one can easily find out which other factors apply and which side won. It is therefore easier to zoom in on relevant cases and to get an initial indication of what a retrieved case means and whether it is relevant. By inspecting the applicable factors of cases, students can judiciously select case squibs for reading. Thus, they can focus on making arguments with cases, not on reading the full opinions of cases that end up not making a contribution to their arguments. We realize that there is value in having students grapple with the complexity of full-text retrieval and complete case opinions. But that is the next step. Practice with CATO is meant to prepare students for this next step. In all likelihood, the development of algorithms for automatically spotting factors in the text of case opinions [23] will open up interesting opportunities to integrate activities with full-text retrieval systems into the instruction with CATO.

8.3. *Second task students practice with CATO: Argumentation guided by examples*

The second main task that students carry out with CATO is an argumentation task, in which the students are given a problem situation and create a written outline of the best arguments that the plaintiff and the defendant can make, supported by cites to cases that they select from CATO’s Case Database. The arguments are supposed to be of comparable complexity as CATO’s Issue-Based Arguments presented earlier. The Workbook instructions provide much guidance. This type of argumentation task is both realistic and important: Much of what legal professionals do is analyzing the legal consequences of a problem and presenting that analysis in writing, in a memo or brief. First-year law students usually practice this kind of task a few times as part of their legal

⁷ There are many different types of full-text retrieval systems (e.g., Boolean, vector space, and probabilistic retrieval). They have in common that a document is considered relevant and is retrieved to the extent that terms mentioned in the query (and perhaps other words based on the same “stem”) appear in the document [59,67].

writing instruction. After that, they may get more practice during their internships but seldom as part of the regular law school curriculum.

As part of the argumentation task, students study argumentation examples, which are generated dynamically by CATO. They use CATO's argumentation examples as models for their own arguments and evaluate their own arguments by comparing them against comparable arguments generated by CATO. Because the examples are generated dynamically, they can be more relevant to the students' on-going work than if pre-stored examples were used. The examples generated by CATO involve the problem situation that the students are analyzing and cases that they may consider citing in their arguments. Also, CATO *tailors* the examples to the students' analysis of the problem, as explained below. Thus, the examples are relevant to the argumentation task at hand, which may heighten the students' interest in the examples and make it easier for them to use the examples as models for their own arguments. Further, CATO annotates the examples to show their underlying structure, revealing argumentation concepts such as relevant similarities, differences, focal abstractions, etc. In this way, CATO *reifies* argument structure that is not visible in a more traditional instructional setting. For example, when a law professor engages a class of students in a Socratic dialog, arguments and counterarguments follow each other in rapid succession. There is usually little time to stop and think about the structure of the arguments. Nor do textbooks on legal argumentation describe in much depth how, for example, a student might make an effective argument that distinctions are or are not significant. The use of CATO may provide a better opportunity for students to reflect upon the structure of arguments, as illustrated by the transcripts of students' explaining argumentation examples generated by CATO, shown in Aleven [1, p. 125ff]. Reification of reasoning structures has been successful in several intelligent tutoring systems [39,62].

The first activity in the argumentation task is to read the facts of the given problem, which is based on the Alabama case *Mason v. Jack Daniel Distillery*, 518 So.2d 130 (Ala.Civ.App.1987). The students identify the factors that apply, using CATO's Case Analyzer tool, which lets them browse information about CATO's set of factors and select the factors they think apply in *Mason*. This tool also provides feedback, comparing the factors selected by the students to its own set of factors for *Mason*. The purpose of this feedback is not to make sure that the students arrive at a "correct" set of factors for *Mason*—we do not think a single correct set exists—but rather to help them in interpreting the meaning of the factors.

As they work on constructing an argument that plaintiff should win in *Mason*, students study argumentation examples generated by CATO in a series of increasingly complex contexts. Initially, they study introductory examples of five basic argument moves: (1) Analogizing a problem to a past case, (2) distinguishing a problem and a past case, (3), (4) downplaying and emphasizing the significance of distinctions, and (5) arguing that weaknesses in a fact situation are not fatal. The students read a general description of each argument move in the Workbooks. They then use CATO's Argument Maker tool to generate examples suggested by the Workbooks. Each example involves a comparison of the current problem situation, *Mason*, and a case that the students select from CATO's Case Database, following the instructions in the Workbook. The examples are based on the students' analysis of *Mason*, that is, the set of applicable factors that they selected, so as to make them maximally relevant to the students' on-going work. CATO annotates

the examples to reveal the underlying structure. As illustrated in Fig. 19, in the top pane of the Argument Maker window, it lists the applicable factors of the problem and the selected case, marking the similarities and distinctions. In the middle pane, CATO shows information about how relevant argumentation concepts apply to the current comparison of cases. For some argument moves, it also displays an argument template. In the bottom pane, finally, CATO displays an argument in English. The students also use the Argument Maker window to engage in a Mini Dialog, in which they practice the skill of distinguishing cases.

In the next phase, the students study argumentation examples in a more complex context, namely, as they search for cases that they can cite in their argument. This task involves important legal research skills that law schools aim to teach. With CATO, students practice a simplified version of the task, as compared to the “open research assignments” which are often part of more traditional legal writing instruction. These assignments typically require that students consult many different resources available in the law library. CATO makes the task more manageable by removing distracting complexity, in much the same way as it makes the students’ theory-testing tasks more manageable. Following the instructions in the Workbook, the students practice two important general legal research strategies, namely, to retrieve (a) cases that are “highly similar” to the problem at hand and (b) cases that they can use to downplay weaknesses. They run queries to the CATO Database to implement these strategies. These queries return cases that are likely to be very useful to support an argument that the plaintiff or the defendant should win in *Mason*, but there is still the need to evaluate the relevance of the retrieved cases in more detail. For starters, the students can look at the applicable factors and the squibs of the retrieved cases, to get an impression of what these cases mean, just as in the theory-testing exercises. The students can also use CATO’s Argument Maker to help them implement a useful strategy for evaluating the relevance of cases, namely, to play out the arguments that could be made with the retrieved cases. At the end of this phase, the students usually have identified a small set of cases that are most relevant to *Mason*.

Finally, students consider CATO’s argumentation examples in the most complex context, namely, as they outline their argument on behalf of the plaintiff in *Mason*, using the cases and other information gathered so far. In this phase, they study a number of Issue-Based Arguments generated by CATO upon request. First, following Workbook instructions, the students identify issues in *Mason* and compare their sets of issues against those identified by CATO, which CATO presents in the form of an Issue-Based Argument that does not cite any cases. Next, the students need to decide how to use the cases they selected previously. They study an example Issue-Based Argument generated by CATO with pre-selected cases, as suggested in one of the Workbooks. Using CATO’s argument as model, the students generate a 1–2 page written argument outline with cites to relevant cases. Finally, students compare their own argument outline to a third Issue-Based Argument. During the evaluation study of CATO, we had planned that each pair of students would compare their argument against one generated by CATO with the same cases that they had cited. Such an argument would have been easy to generate using CATO and would have made for a particularly relevant comparison. But time permitted us only to provide a handout with arguments generated by CATO that was the same for each student.

Nonetheless, it is an advantage of dynamic example generation that such a comparison is possible.

In short, during the argumentation task, students studied a significant number of argumentation examples. The examples were relevant to the student's task at hand and, in a simple way, were tailored to their analysis of the problem. This tailoring made it easier for the students to use the examples as models for their own argument and to compare their work against the examples. Further, the examples made explicit the underlying structure of the arguments. During the CATO instruction, students have ample opportunity to study and reflect on the reified structure, a significant advantage afforded by a computer-based learning environment equipped with an argumentation model.

9. Evaluation study 2: Instructional effectiveness of CATO

We conducted an evaluation study to assess the effectiveness of the CATO instruction in comparison to the best way of teaching the same basic argumentation skills by traditional methods. In fact, as we will see, the control group instruction in this experiment represents a more ideal instructional situation than that typically found in American law school curricula. The experiment also provided an opportunity to evaluate CATO's Issue-Based Arguments by comparing them against those of the best students who participated.

9.1. Subjects

The study was carried out in the context of a second-semester legal writing course at the University of Pittsburgh School of Law. The subjects in the experiment were first-year law students, recruited from those taking the course. Participation in the experiment was recommended by the course instructor but was voluntary. Those students who elected to participate were assigned randomly to an experimental group of 16 students and a control group of 14 students. The students who elected not to participate in the experiment attended the same instructional sessions as the control group. Each group was divided into 3 sections of roughly equal size. The average class size in the control group was 8 students.

9.2. Procedure

The students in the experimental group worked through the CATO curriculum, which consisted of the Casebook and Workbooks described previously. During a three-week period, instead of attending their regular class meetings, they went to a CATO computer lab in the law school set up specifically for the experiment. Each student worked with CATO for nine 50-minute sessions for a total of 7.5 hours of instruction. After an introductory classroom session, during which a human instructor (Kevin Ashley) introduced students to CATO, factors, and Word Perfect, students collaborating in pairs used the CATO program to work through theory-testing and argumentation problems, as described in the previous sections.

The control group instruction was designed to teach the same basic skills of making arguments with cases in a traditional way, in the same amount of time as that covered by the CATO instruction. The control group instruction was based on the same casebook and was taught by an instructor who was not familiar with the CATO model. During four classroom sessions, the course instructor used a Socratic method to present a framework of inquiry for trade secret law, synthesized from the casebook cases. During two moot court sessions led by the instructor, the students made oral arguments about two problem situations. The instructor, combining the roles of “judge” and teacher, moderated the argument exchange. The students prepared for these sessions outside of class, spending at least 75 minutes each time.

9.3. Tests

To assess the improvement in students’ argumentation skills, we administered a *Basic Argument Skills* pre-test and post-test, both in-class exams. Each test consisted of argumentation problems, involving a problem situation and three cases, for which squibs were presented. We also administered a second post-test designed to assess whether students are able to transfer and use basic argumentation skills in a more complex task, a *Legal Memo-Writing Assignment*. The students were given a problem situation and six case squibs and were asked to produce a six-page memo, in which they stated the plaintiff’s and the defendant’s strongest argument on the issues that they thought were raised by the problem. Students had one week to complete this assignment. The tests and grading criteria were created in consultation with the legal writing instructor who taught the control group. All tests were graded in a blind test by the legal writing instructor. Without informing him, we included in the materials to be graded a set of answers generated by the CATO program, hand-written and formatted in a way so as to disguise the fact that they had been computer-generated, but otherwise exactly as CATO had generated them. CATO’s answers to each of these tests contained one or more Issue-Based Arguments.

9.4. Results

The results, presented in Table 2, indicate that on the Basic Argument Skills Tests, both groups’ scores improved from pre-test to post-test. The letter grades were computed by converting the numerical grades. In each group, the improvement was statistically significant (experimental group: $t(15) = -3.4$, $p = 0.004$; control group: $t(13) = -3.7$, $p = 0.002$). There was no significant difference between the groups in pre-test, post-test,

Table 2
Results of the basic argument skills tests and legal memo writing assignment

	Basic argument skills				Memo writing		
	Pre-test		Post-test		Prev	Post-test	
CATO group	60	C–	70	C+	63	70	B–
Control group	55	D	68	C	63	79	B+
CATO’s arguments	81	B+	87	A–		62	C

or gain scores (pre-test: $t(28) = 1.56$, $p = 0.13$; post-test: $t(28) = 0.57$, $p = 0.58$; gain scores: $t(28) = -0.66$, $p = 0.51$). CATO's answers ranked among the best on the Basic Argument Skills tests (see Table 2). CATO came in first on the pre-test and third on the post-test, or fourth and sixth respectively if one focuses only on the first question, which more students finished than the second.

On the Memo-Writing Assignment, the control group scores were higher than the scores in the experimental group. The difference was statistically significant ($t(25) = -2.38$, $p = 0.03$). No statistically significant difference existed between the groups on the writing assignment of the previous semester ($t(25) = 0.05$, $p = 0.96$). CATO's memo did not receive a high grade.

9.5. Discussion

The results on the Basic Argument Skills Tests indicate that the CATO instruction leads to a statistically significant improvement of students' basic skills of making arguments with cases, comparable to that achieved by an experienced legal writing instructor teaching small groups of students. This result should be interpreted in light of the fact that the control group instruction represents a very high standard of comparison. First, the subjects were among the more difficult students to teach. They were enrolled in a program that provides extra-intensive legal writing instruction. They had been selected for participation in this program, by the law school's admissions committee, on the basis of various indicators of law school aptitude suggesting a need for special instruction. Second, the control group instruction involved very small groups: 8 students on average. It therefore represents a far more ideal situation than one normally finds in legal writing classes. Finally, the control group instructor has excellent credentials. He is the director of the legal writing program at the University of Pittsburgh School of Law as well as the special program described above. Experienced in teaching such students, he is a dedicated teacher who enjoys an excellent rapport with his students, as evidenced for example by the students' enthusiasm for the oral argument sessions. Seen in this light, the result that the instruction with CATO was equally effective as the control group instruction is impressive.

On the other hand, the results on the Legal Memo-Writing Assignment show the limits of CATO's effectiveness. The Legal Memo-Writing Assignment was a transfer task, going well beyond what students were asked to do during the experimental or control group instruction. It posed a more advanced and complex task than the Basic Argument Skills tests or the argumentation tasks students carried out with CATO. It involved more cases and posed additional problems of selecting the best argument and composition. The combined results on the Basic Argument Skills test and the Legal Memo-Writing Assignment suggest that while both the experimental and control group learned basic skills of making arguments with cases, the control group was better able to integrate these skills in a more complex context.

In part, this difference may be due to the differences in instructional approach between the two conditions. As mentioned, CATO teaches in a component-wise, example-based manner and suppresses some distracting complexity. During the CATO instruction, complexity was introduced gradually. Basic Argument Moves were addressed early on, the more complex Issue-Based Arguments not until the last couple of sessions. Unfortunately,

most students ran out of time while working on the more complex topics. The control group instruction on the other hand did not reduce complexity and did not follow a component-based approach. As is common in American law schools, the legal writing instructor who taught the control group instruction taught the argumentation skills in a “holistic” way. He discussed real cases and engaged students in arguments about these cases. He did not discuss any methods for using cases in arguments.

The results of the experiment indicate that a component-wise, example-based approach can be effective in helping students to learn basic argumentation skills. In this approach there is time and opportunity for students to see and reflect on argument structure. At the same time, a limitation of this approach seems to be that it provides less help than would be ideal with the application of basic argumentation skills in a more complex context. Thus, there is a need to make the instruction with CATO more “integrated”.

The comments that the instructor wrote on CATO’s memo provide us with much information about CATO’s Issue-Based Arguments. As we had found in earlier experiments [4], CATO’s answers on the Basic Argument Skills Tests are on par with those of the best students. As mentioned, CATO’s answers to these tests comprised Issue-Based Arguments as well as basic argument moves in isolation, although no arguments about the significance of distinctions. CATO’s good grades on the Basic Argument Skills pre-test and post-test provide some evidence that the Factor Hierarchy and CATO’s Issue-Based are successful on small problems. Interestingly, the legal writing instructor said afterwards that he had not been aware that one of the exams had been computer-generated or was somehow different from the rest.

Although CATO’s Legal Memo-Writing Assignment did not receive a high grade, reflecting the tougher standards used to grade the assignment, the legal writing instructor did have a number of good things to say about CATO’s memo. Although this time he did not think CATO’s arguments had been produced by one of his students, he was impressed when it was revealed to him that the memo had been produced by a computer program (see [1, p. 175]). He stated that CATO’s memo “did a very good job as far as thoroughness and accuracy [are concerned].” He also said that he had used CATO’s memo to some extent to guide his grading: “It was like to me a high-level outline of what was going on in those cases.”

On the other hand, the legal writing instructor also stated that the analysis was “too fragmented.” He found many statements “vague” or “conclusory” and often asked for elaboration. He found that the argument was thorough but in a way that lacked judgment as to what was most important. Also, he insisted on a different format for case-citing arguments (“discuss case law first and then apply to our facts”) and a different top-level organization. Clearly, these critiques imply that CATO’s Issue-Based Arguments were less than ideal models for students to prepare for the Legal Memo-Writing Assignment.

In sum, as a component of the CATO program’s instruction, the Factor Hierarchy helped achieve good results in teaching basic argumentation skills. Further, the legal writing instructor’s comments that CATO’s argument was “accurate” and served for him as a “high-level outline” confirm that the links of the Factor Hierarchy were relevant and accurate. Many criticisms of CATO’s Issue-Based Arguments, such as a wrong choice of top-level organization and of “fragmented discussion”, seem to be due mostly to a lack of sophistication in its text-generation capabilities. But some criticisms also seem to point to

limitations of the Factor Hierarchy itself. More of a synthesis of the various cases may be required. Also, it may be necessary to discuss case facts in more detail. For this purpose, the Factor hierarchy's evidential links are important but may not be sufficient.

The two needs identified above, (a) for more holistic, integrated instruction, and (b) for case synthesis, can to a large degree be addressed together. Issues, theory testing, and argumentation should be integrated more tightly in the CATO instruction. There are, in principle, important links between identifying issues, testing theories, and constructing and organizing arguments. Theories induced from cases, such as the theories addressed in the CATO instruction, can be cited as warrants in arguments about how a problem situation should be decided. This link should be made more explicit during the CATO instruction. The instruction should be organized so that students test theories in the context of an argumentation task. From the start, the students should be involved in identifying issues in a problem, identifying strengths and weaknesses related to each issue, formulating alternative theories about whether and why those strengths outweigh the weaknesses, and testing these theories by running queries. The CATO instruction should teach them to present a theory and supporting cases in their arguments, producing the kind of case synthesis called for by the legal writing instructor.

In order to support this kind of instruction, CATO's Issue-Based Arguments would need to be changed so that the discussion of an issue is centered around a theory. The Factor Hierarchy will be instrumental in representing theories and their links with issues. Importantly, the Factor Hierarchy would help in exploring theories at different levels of abstraction, which is a desirable property [20]. The Factor Hierarchy could also be used to suggest variations of a theory that are stronger or weaker, by adding or removing relevant factors.⁸

It will also be very useful to explore ways of combining instruction with CATO and more holistic instruction. For example, a human instructor could demonstrate how to carry out theory-testing or argumentation tasks with CATO in front of a class of students, using a data projector to project CATO's output onto a big screen. In doing so, the instructor could bring up and illustrate a rich set of connections between the components of CATO's argumentation model or cases retrieved from its database, going beyond the connections that CATO itself is capable of making. A skilled instructor could also make connections with knowledge about substantive law, arguments, and argumentation techniques not represented in CATO.⁹ Following such classroom sessions, the students could use CATO in

⁸ A recent paper formalized notions of reasoning with factors, cases, values, and theories, viewing case-based reasoning as a process of searching a space of factor-based theories [16].

⁹ For example, CATO currently does not represent the policies underlying trade secrets law. Nonetheless, it is a suitable vehicle to bring up and illustrate some of these policies. During two classroom sessions unrelated to the experiment described in the current paper, a law professor (Kevin Ashley) used CATO queries to illustrate two of the policies underlying trade secrets law and to illustrate how they differ from the policies underlying other areas of intellectual property law, such as patent law. (CATO's output was projected onto a projection screen by means of a data projector.) The requirement that information must be substantially secret in order to be protected as a trade secret can be illustrated for example by running a query for cases with factor F20 Info-Known-To-Competitors. This query returns only cases in which the defendant won (at least it did so at the time). The fact that, apparently, plaintiffs in trade secrets cases cannot win if the alleged trade secret is known outside their business illustrates a difference with patent law, since patented information is in the public domain. A second

the lab to solidify their understanding. It may be effective also to use CATO in the context of moot court sessions, given that the control group moot court sessions generated great enthusiasm. For example, students could use CATO to prepare their oral arguments.¹⁰

10. Conclusion

Researchers in the field of AI and Law have long been interested in modeling the ways in which attorneys use cases in arguments. We have presented a model of case-based legal argumentation, the CATO model, which focuses on the representation of middle-level normative background knowledge and the use of this knowledge for purposes of case selection and argumentation. We have described and illustrated a Factor Hierarchy, a novel knowledge source implemented in the CATO program which represents knowledge about the meaning of the factors that are used to represent cases. We have described and illustrated novel methods for using background knowledge (1) to organize multi-case arguments by issues, (2) to generate context-sensitive arguments about the similarity of cases, which focus on alternative characterizations of a distinction's significance, and (3) to assess the similarity of cases in order to select the best cases to cite in an argument. We have described how the model can be used as a central component of an effective intelligent learning environment designed to provide instruction in basic argumentation skills.

The first evaluation study presented in this paper provides evidence that CATO's arguments about the significance of distinctions are sound. The results show that, compared to standard CBR methods that do not rely on background knowledge, CBR methods that apply background knowledge to reason about the significance of distinctions help to make better predictions of the outcome of trade secrets cases. They make it possible to assess the similarity of cases in a more discriminating manner. The results of the

CATO query illustrated a second difference between the policies underlying the two areas of the law. The fact that trade secrets law does not aim to protect owners of secret information against competitors who develop that information independently can be illustrated by running a query for cases with factor F17 Info-Independently-Generated. That query returns only cases won by the defendant. Patent law on the other hand does protect against independent re-invention. This kind of demonstration with the CATO program is powerful and illustrates to beginning law students the importance of cases in legal reasoning. Using a standard information retrieval system, it would not have been feasible to carry out this kind of exercise, for the same reasons that theory-testing exercises would be much harder to do with a standard legal information retrieval system. The exercise would have bogged down in the complexity of trying to figure out what the retrieved cases mean.

A proposal for integrating reasoning about the policies or purposes underlying trade secrets law into the HYPO/CATO model was presented by Berman and Hafner [17]. Since this critique was published, a computational model [51] and a theoretical account [16] of relations between principles and cases have been published.

¹⁰ During one of the classroom sessions mentioned above, we saw that CATO can help generate a considerable amount of excitement among beginning law students when used in a context that resembles a moot court. CATO was used to explore a legal issue that the students had been hotly debating as part of an assignment in a related class. The issue involved the distinction between an employee and an independent contractor, an area for which CATO has a small Case Database. Guided by the professor, the students formulated CATO queries to test some of the theories they had debated with their peers. They cheered as their queries turned up cases confirming their theories and turned up no counterexamples. Again, this exercise would not have been practical using a standard legal information retrieval system.

experiment suggest, although they do not prove conclusively, that the same improvement could not have been obtained with simpler CBR methods that take into account whether distinctions between cases exist, but do not reason about their significance and do not apply background knowledge. The predictions based on the CBR methods were not as good as those obtained by one of the inductive learning methods that was tested, Naïve Bayes. This result may to some degree be attributed to the fact that Naïve Bayes is not required to provide arguments to support its decisions and thus has more freedom in how it uses the available information to make predictions. But the result suggests also that it is worthwhile to continue to look for better ways of applying normative background knowledge within a CBR framework. Regardless, it is a significant outcome that the use of background knowledge in CBR methods leads to better predictions, compared to CBR methods that do not apply background knowledge. Apparently, CATO's arguments about the significance of distinctions bear some relation to the way legal cases are decided, which inspires confidence that these arguments are relevant and sound. Further, the results of this evaluation study provide strong support for the claim that middle-level normative background knowledge is an important component of case-based legal reasoning.

The second evaluation study provided evidence for the effectiveness of instruction with the CATO program. We compared instruction with CATO to traditional legal writing instruction given to very small groups of students led by an experienced legal writing instructor. The results show that the students who worked with the CATO program showed comparable learning gains as students who were instructed in a more traditional manner. The experiment also showed that on a transfer task, a more complex legal writing assignment, the students in the control group performed better than students who had worked with CATO. These results indicate that in the legal domain, an intelligent learning environment that provides example-based component-wise instruction is effective in teaching basic argumentation skills but that more is needed to help students to transfer these skills to a more complex context. The experiment also provides some evidence that CATO's Factor Hierarchy is sound. Arguments generated by CATO were deemed (by the control group instructor) to be about as good as those of the best students, on tests designed to measure basic argumentation skills. However, when graded according to the tougher standards applied to the transfer task, CATO's arguments did not fare so well. In order to be more effective, the instruction with CATO must be re-organized to achieve a better integration of argumentation, theory-testing, and case synthesis. We have described in some detail how this integration might be achieved.

The research on CATO makes contributions to both the theoretical and the practical research goals of the field of AI and Law. CATO's theoretical contribution is a demonstration that in areas of the law where cases are naturally represented in terms of factors (and we have argued that there are many such areas), middle-level normative background knowledge about the meaning of those factors is a key ingredient of case-based legal argumentation. The examples, methods, and evaluation results presented in the paper confirm and detail the central role that this type of knowledge plays in such frequently-occurring argumentation tasks as organizing arguments by issues and evaluating the significance of distinctions. It has even been argued that CATO's use of background knowledge is a step towards generating arguments based on the policies and purposes underlying the given area of the law [6]; see also [17]. The practical contribution of the

current work to the research field of AI and Law is a demonstration that a computational model of case-based legal argumentation can be used effectively to provide computer-based instruction to beginning law students. CATO is not the first intelligent tutoring system for the legal domain, but it is unique in the extent to which it focuses on skills of argumentation and reasoning with cases.

The novel methods for reasoning about the significance of differences presented in this paper constitute a contribution to the research field of CBR as well. Similarity assessment and explaining similarity are important concerns in CBR research. The current research demonstrates novel ways in which (fairly weak) background knowledge can be used in a context-sensitive manner for similarity assessment. CATO's techniques are appropriate:

- in adversarial domains, where arguers justify conclusions about how a problem should be classified by comparing it to past cases, and
- where abstract terms are available that bear on the overall decision or classification of a problem and inform the significance of base-level similarities and differences of cases, and
- where there is freedom in interpreting cases, so that an arguer can select from among alternative (abstract) interpretations of cases based at least in part on the outcome being argued for.

The main difference with existing CBR approaches to explaining similarity is that CATO supports the exploration of opposite viewpoints of the significance of distinctions, by characterizing and re-characterizing their meaning in more abstract terms. This kind of dialectic approach, in which alternative viewpoints on the similarity of given cases are explored, may have a place in other CBR applications as well. Unless a CBR program has a means of checking the solutions that it proposes (and few CBR programs do), it makes sense that it would look also at the plausibility of alternative solutions based on alternative cases. For example, a program for estimating construction costs (for example, see [69]) might offer best-case and worst-case scenarios.

Finally, CATO's main contribution to the field of AI and Education is a demonstration that basic argumentation skills can be taught effectively in a component-wise, example-based manner, by an instructional environment that employs a computational model of case-based argumentation.

Acknowledgements

This paper is based on the author's doctoral dissertation work, which was carried out while he was at the University of Pittsburgh. The author wishes to thank Kevin Ashley for being a terrific mentor and Bruce Buchanan, Alan Lesgold, and Kurt VanLehn for their constructive and insightful guidance. Bruce Buchanan gave advice on the follow-up study reported in this paper. Stefanie Brüninghaus provided advice and assistance related to the evaluation studies reported in the paper and kindly provided new cases for CATO's Case Database. The anonymous reviewers and the special issue editors also provided valuable input. Any remaining inaccuracies are the author's responsibility.

The author has been supported by a Mellon Fellowship. The research has been sponsored by grants from the National Science Foundation, West Publishing Company, Digital Equipment Corporation, Tektronix, the National Center for Automated Information Research, and the University of Pittsburgh ECAC Advanced Instructional Technology Program. Their support is gratefully acknowledged.

References

- [1] V. Alevén, Teaching case-based argumentation through a model and examples, Unpublished Ph.D. Dissertation, Intelligent Systems Program, University of Pittsburgh, Pittsburgh, PA, 1997, <http://www.cs.cmu.edu/~aleven>.
- [2] V. Alevén, K.D. Ashley, Evaluating a learning environment for case-based argumentation skills, in: Proc. 6th International Conference on Artificial Intelligence and Law, ICAIL-97, Melbourne, Australia, ACM, New York, 1997, pp. 170–179.
- [3] V. Alevén, K.D. Ashley, How different is different? Arguing about the significance of similarities and differences, in: I. Smith, B. Faltings (Eds.), *Advances in Case-Based Reasoning: Proc. 3rd European Workshop, EWCBR-96*, in: *Lecture Notes in Artificial Intelligence*, Vol. 1168, Springer, Berlin, 1996, pp. 1–15.
- [4] V. Alevén, K.D. Ashley, Using a well-structured model to teach in an ill-structured domain, in: Proc. 17th Annual Conference of the Cognitive Science Society, Lawrence Erlbaum, Mahwah, NJ, 1995, pp. 419–424.
- [5] J.R. Anderson, A.T. Corbett, K.R. Koedinger, R. Pelletier, Cognitive tutors: Lessons learned, *J. Learning Sci.* 4 (1995) 167–207.
- [6] K.D. Ashley, An AI model of case-based legal argument from a jurisprudential viewpoint, *Artif. Intell. Law* 10 (2002) 163–218.
- [7] K.D. Ashley, Reasoning with cases and hypotheticals in HYPO, *Internat. J. Man-Machine Stud.* 34 (6) (1991) 753–796.
- [8] K.D. Ashley, *Modeling Legal Argument: Reasoning with Cases and Hypotheticals*, MIT Press/Bradford Books, Cambridge, MA, 1990.
- [9] K.D. Ashley, V. Alevén, Reasoning symbolically about partially matched cases, in: M. Pollack (Ed.), *Proc. IJCAI-97*, Nagoya, Japan, Morgan Kaufmann, San Francisco, CA, 1997, pp. 335–341.
- [10] K.D. Ashley, V. Alevén, Generating dialectical examples automatically, in: Proc. AAAI-92, San Jose, CA, AAAI Press, Menlo Park, CA, 1992, pp. 654–660.
- [11] K.D. Ashley, R. Desai, J.M. Levine, Teaching case-based argumentation concepts using dialectic arguments vs. didactic explanations, in: S.A. Cerri, G. Gouardères, F. Paraguaçu (Eds.), *Proc. 6th International Conference on Intelligent Tutoring Systems, ITS 2002*, Springer, Berlin, 2002, pp. 585–595.
- [12] K.D. Ashley, B.M. McLaren, Reasoning with reasons in case-based comparisons, in: M. Veloso, A. Aamodt (Eds.), *Case-Based Reasoning Research and Development, Proc. 1st International Conference on Case-Based Reasoning, ICCBR-95*, in: *Lecture Notes in Artificial Intelligence*, Vol. 1010, Springer, Berlin, 1995, pp. 133–144.
- [13] K.D. Ashley, E.L. Rissland, Waiting on weighting: A symbolic least commitment approach, in: Proc. AAAI-88, St. Paul, MN, AAAI Press/MIT Press, Cambridge, MA, 1988, pp. 239–244.
- [14] R. Bareiss, B.W. Porter, K.W. Murray, Supporting start-to-finish development of knowledge bases, *Machine Learning* 4 (1989) 259–283.
- [15] R. Bergmann, W. Wilke, On the role of abstraction in case-based reasoning, in: Proc. European Workshop on Case-Based Reasoning, Lausanne, Switzerland, 1996, pp. 28–43.
- [16] T. Bench-Capon, G. Sartor, Theory-based explanation of case law domains, in: Proc. 8th International Conference on Artificial Intelligence and Law, St. Louis, MO, ACM, New York, 2001, pp. 12–21.
- [17] D.H. Berman, C.D. Hafner, Representing teleological structure in case-based legal reasoning: The missing link, in: Proc. 4th International Conference on Artificial Intelligence and Law, Amsterdam, ACM, New York, 1993, pp. 50–59.

- [18] D.H. Berman, C.D. Hafner, Incorporating procedural context into a model of case-based legal reasoning, in: *Proc. 4th International Conference on Artificial Intelligence and Law*, Oxford, ACM, New York, 1991, pp. 12–20.
- [19] L.K. Branting, Stratified case-based reasoning in non-refinable abstraction hierarchies, in: D.B. Leake, E. Plaza (Eds.), *Case-Based Reasoning Research and Development*, Second International Conference on Case-Based Reasoning, ICCBR-97, Springer, Berlin, 1997, pp. 519–530.
- [20] L.K. Branting, A computational model of ratio decidendi, *Artif. Intell. Law* 2 (1994) 1–31.
- [21] L.K. Branting, Building explanations from rules and structured cases, *Internat. J. Man-Machine Stud.* 34 (6) (1991) 797–838.
- [22] L.K. Branting, D.W. Aha, Stratified case-based reasoning: Reusing hierarchical problem-solving episodes, in: *Proc. IJCAI-95*, Montreal, Quebec, Morgan Kaufmann, San Mateo, CA, 1995, pp. 384–390.
- [23] S. Brüninghaus, K.D. Ashley, Improving the representation of legal case texts with information extraction methods, in: *Proc. 8th International Conference on Artificial Intelligence and Law*, St. Louis, MO, ACM, New York, 2001, pp. 42–51.
- [24] S.J. Burton, *An Introduction to Law and Legal Reasoning*, Montreal, Quebec, 2nd Edition, Little, Brown and Company, Boston, MA, 1995.
- [25] R. Burrow, G. Weber, Example explanation in learning environments, in: C. Frasson, G. Gauthier, A. Lesgold (Eds.), *Intelligent Tutoring Systems*, 3rd International Conference, ITS'96, Springer, Berlin, 1996, pp. 457–465.
- [26] F. Centinía, T. Routen, A. Hartmann, C. Hegarty, STATUTOR: Too intelligent by half?, in: J.C. Hage, T.J.M. Bench-Capon, M.J. Cohen, H.J. van den Herik (Eds.), *Legal Knowledge Based Systems JURIX'95: Telecommunication and AI & Law*, Koninklijke Vermande, Lelystad, Netherlands, 1995, pp. 121–132.
- [27] Y.S. Chee, S. Xu, SIPLeS: Supporting intermediate smalltalk programming through goal-based learning scenarios, in: B. du Boulay, R. Mizoguchi (Eds.), *Artificial Intelligence in Education*, Proceedings of AI-ED 97 World Conference, IOS Press, Amsterdam, 1997, pp. 95–102.
- [28] M. Chi, M. Bassok, M. Lewis, P. Reimann, R. Glaser, Self-explanations: How students study and use examples in learning to solve problems, *Cognitive Sci.* 15 (1989) 145–182.
- [29] C. Conati, K. VanLehn, Toward computer-based support of meta-cognitive skills: A computational framework to coach self-explanation, *Internat. J. Artif. Intell. Education* 11 (2000).
- [30] B. Cuthill, R. McCartney, Issue spotting in CHASER, *Artif. Intell. Law* 2 (1994) 83–111.
- [31] T. Dietterich, Statistical tests for comparing supervised classification learning algorithms, Technical Report, Department of Computer Science, Oregon State University, Corvallis, OR, 1996, <http://citeseer.nj.nec.com/dietterich96statistical.html>.
- [32] D.C. Edelson, Learning from cases and questions: The Socratic case-based teaching architecture, *J. Learning Sci.* 5 (4) (1996) 357–410.
- [33] W. Ferguson, R. Bareiss, L. Birnbaum, R. Osgood, ASK systems: An approach to the realization of story-based teachers, *J. Learning Sci.* 2 (1) (1992) 95–134.
- [34] P.W. Fung, A. Adam, Conceptual graphs for similarity measurement in a case-based physics problem solver, in: J. Greer (Ed.), *Proceedings of AI-ED 95, the 7th World Conference on Artificial Intelligence in Education*, Association for the Advancement of Computing in Education (AACE), Charlottesville, VA, 1995, pp. 170–177.
- [35] A. Gardner, *An Artificial Intelligence Approach to Legal Reasoning*, MIT Press, Cambridge, MA, 1987.
- [36] K.M. Gupta, Taxonomic conversational case-based reasoning, in: D.W. Aha, I. Watson (Eds.), *Case-Based Reasoning Research and Development*, 4th International Conference on Case-Based Reasoning, ICCBR 2001, in: *Lecture Notes in Computer Science*, Vol. 2080, Springer, Berlin, 2001, pp. 219–233.
- [37] C. Joiron, D. Leclerc, A case base model for case based forum: Experimentation on pediatric pain management, in: J.D. Moore, C.L. Redfield, W.L. Johnson (Eds.), *Artificial Intelligence in Education: AI-ED in the Wired and Wireless Future*, Proceedings of AI-ED 2001, IOS Press, Amsterdam, 2001, pp. 111–121.
- [38] A. Kass, R. Burke, E. Blevis, M. Williamson, Constructing learning environments for complex social skills, *J. Learning Sci.* 3 (4) (1993) 387–427.
- [39] K.R. Koedinger, J.R. Anderson, Reifying implicit planning in geometry: Guidelines for model-based intelligent tutoring system design, in: S. Lajoie, S. Derry (Eds.), *Computers as Cognitive Tools*, Erlbaum, Mahwah, NJ, 1993, pp. 15–45.
- [40] J.L. Kolodner, *Case-Based Reasoning*, Morgan Kaufmann, San Mateo, CA, 1993.

- [41] P. Koton, Reasoning about evidence in causal explanations, in: Proc. AAAI-88, St. Paul, MN, AAAI Press/MIT Press, Cambridge, MA, 1988, pp. 256–261.
- [42] D.B. Leake (Ed.), Case-Based Reasoning: Experiences, Lessons, and Future Directions, AAAI Press/MIT Press, Menlo Park, CA, 1996.
- [43] D.B. Leake, An indexing vocabulary for case-based explanation, in: Proc. AAAI-91, Anaheim, CA, AAAI Press/MIT Press, Cambridge, MA, 1991, pp. 10–15.
- [44] E.H. Levi, *An Introduction to Legal Reasoning*, University of Chicago Press, Chicago, IL, 1949.
- [45] K.N. Llewellyn, *The Bramble Bush. On Our Law and Its Study*, Oceana Publications, New York, 1930.
- [46] B.M. McLaren, K.D. Ashley, Case-based comparative evaluation in TRUTH-TELLER, in: Proc. 17th Annual Conference of the Cognitive Science Society, Lawrence Erlbaum, Mahwah, NJ, 1995, pp. 72–77.
- [47] E. Melis, Solution-relevant abstractions constrain retrieval and adaptation, in: D.B. Leake, E. Plaza (Eds.), *Cased-Based Reasoning Research and Development*, Second International Conference on Case-Based Reasoning, ICCBR-97, Springer, Berlin, 1997, pp. 383–392.
- [48] T.M. Mitchell, *Machine Learning*, McGraw-Hill, New York, 1997.
- [49] A.J. Muntjewerff, J.A. Breuker, Evaluating PROSA, a system to train solving legal cases, in: J.D. Moore, C.L. Redfield, W.L. Johnson (Eds.), *Artificial Intelligence in Education: AI-ED in the Wired and Wireless Future*, Proceedings of AI-ED 2001, IOS Press, Amsterdam, 2001, pp. 278–285.
- [50] R.K. Neumann Jr., *Legal Reasoning and Legal Writing; Structure, Style, and Strategy*, Little, Brown and Company, Boston, MA, 1990.
- [51] K. Nitta, M. Shibasaki, T. Sakata, T. Yamaji, W. Xianchang, H. Ohsaki, S. Tojo, I. Kokubo, New-HELIC-II: A software tool for legal reasoning, in: Proc. 5th International Conference on Artificial Intelligence and Law, College Park, MD, ACM, New York, 1995, pp. 287–296.
- [52] E.L. Rissland, Dimension-based analysis of hypotheticals from supreme court oral argument, in: Proc. 2nd International Conference on Artificial Intelligence and Law, Vancouver, BC, ACM, New York, 1989, pp. 111–120.
- [53] E.L. Rissland, K.D. Ashley, A note on dimensions and factors, *Artif. Intell. Law* 10 (2002) 65–77.
- [54] E.L. Rissland, D.B. Skalak, CABARET: Rule interpretation in a hybrid architecture, *Internat. J. Man-Machine Stud.* 34 (6) (1991) 839–887.
- [55] E.L. Rissland, D.B. Skalak, M.T. Friedman, BankXX: Supporting legal arguments through case-based search, *Artif. Intell. Law* 4 (1996) 1–71.
- [56] E.L. Rissland, E.M. Valcarce, K.D. Ashley, Explaining and arguing with examples, in: Proc. AAAI-84, Austin, TX, AAAI Press, Menlo Park, CA, 1984, pp. 288–294.
- [57] B. Roth, A typology of moves involved in case comparison, in: Proc. 8th International Conference on Artificial Intelligence and Law, St. Louis, MO, ACM, New York, 2001, pp. 129–138.
- [58] T. Routen, Reusing formalisations of legislation in a tutoring system, *Artif. Intell. Rev.* 6 (1992) 145–159.
- [59] G. Salton, M.J. McGill, *Introduction to Modern Information Retrieval*, McGraw-Hill, New York, 1983.
- [60] S.L. Salzberg, On comparing classifiers: Pitfalls to avoid and a recommended approach, *Data Mining Knowl. Discov.* 1 (1997) 317–328.
- [61] R.C. Schank, Case-based teaching: Four experiences in educational software design, *Interactive Learning Environ.* 1 (4) (1990) 231–253.
- [62] L. Schauble, K. Raghavan, R. Glaser, The discovery and reflection notation: A graphical trace for supporting self-regulation in computer-based laboratories, in: S. Lajoie, S. Derry (Eds.), *Computers as Cognitive Tools*, Erlbaum, Mahwah, NJ, 1993, pp. 319–337.
- [63] T.J. Schult, P. Reimann, Dynamic case-based tutoring: A cognitive science approach, in: J. Greer (Ed.), *Proceedings of AI-ED 95, the 7th World Conference on Artificial Intelligence in Education*, Association for the Advancement of Computing in Education (AACE), Charlottesville, VA, 1995, pp. 154–161.
- [64] D.M. Sherman, Expert systems and ICAI in tax law: Killing two birds with one AI stone, in: Proc. 2nd International Conference on Artificial Intelligence and Law, Vancouver, BC, ACM, New York, 1989, pp. 74–80.
- [65] B. Smyth, M.T. Keane, Adaptation-guided retrieval: Questioning the similarity assumption in reasoning, *Artificial Intelligence* 102 (1998) 249–293.
- [66] G. Span, LITES, an intelligent tutoring system for legal problem-solving in the domain of Dutch Civil Law, in: Proc. 4th International Conference on Artificial Intelligence and Law, Amsterdam, ACM, New York, 1993, pp. 76–81.

- [67] H.R. Turtle, Text retrieval in the legal world, *Artif. Intell. Law* 2 (1) (1995) 5–54.
- [68] H. Watanabe, M. Arai, S. Takei, Case-based evaluation of novice programs, in: J.D. Moore, C.L. Redfield, W.L. Johnson (Eds.), *Artificial Intelligence in Education: AI-ED in the Wired and Wireless Future*, Proceedings of AI-ED 2001, IOS Press, Amsterdam, 2001, pp. 55–64.
- [69] I. Watson, S. Perera, The evaluation of a hierarchical case representation using context-guided retrieval, in: D.B. Leake, E. Plaza (Eds.), *Cased-Based Reasoning Research and Development*, Second International Conference on Case-Based Reasoning, ICCBR-97, in: *Lecture Notes in Computer Science*, Vol. 1266, Springer, Berlin, 1997, pp. 255–266.
- [70] G. Weber, Episodic learner modeling, *Cognitive Sci.* 20 (1996) 195–236.
- [71] E. Wenger, *Artificial Intelligence and Tutoring Systems: Computational and Cognitive Approaches to the Communication of Knowledge*, Morgan Kaufmann, Los Altos, CA, 1987.
- [72] D.M. Zhang, L. Alem, Using case-based reasoning for exercise design in simulation-based training, in: C. Frasson, G. Gauthier, A. Lesgold (Eds.), *Intelligent Tutoring Systems*, Third International Conference, ITS'96, Springer, Berlin, 1996, pp. 560–568.