An Agenda for Empirical Research in AI and Law

L. Karl Branting
LiveWire Logic, Inc.
2700 Gateway Centre Blvd.
Morrisville, NC 27606
karl.branting@livewirelogic.com

ABSTRACT

Market forces have fueled a rapid growth in practical systems for legal problem solving. Unfortunately, the AI and Law research community has been relatively disengaged from this process. This paper argues that the AI and Law community could make a larger contribution to practical legal system development by focusing more on task analysis and empirical validation to insure that computational and formal models correspond to actual problem-solving behavior. In particular, the absence of task and corpus analysis has led to models of precedent-based legal reasoning that are inconsistent with actual problem-solving behavior in the Anglo-American legal system.

1. INTRODUCTION

In recent years there has been a rapid growth in the number of commercial and governmental systems for legal problem solving in active use. Four factors have contributed to the growing automation of legal services. First, there is a substantial unmet demand for legal services by the public, termed the "latent market" by Richard Susskind [Susskind, 1998]. The high cost of attorney's fees means that individuals frequently are unable to afford answers to basic legal questions. These costs are exacerbated by uncertainty in evolving areas of law (e.g. consumer protection), delays in litigation, and lack of predictability [Berman and Hafner, 1989].

A second factor is that the World Wide Web constitutes an electronic infrastructure for the electronic distribution of legal services. The WWW provides a uniform computer interface familiar to increasing numbers of both litigants and attorneys. It largely eliminates the problems of software distribution, hardware inconsistencies, and unfamiliar interfaces that were obstacles to acceptance of automated legal reasoning systems in the 1980s [Morrison, 1989].

A third factor is that legal expert systems are increasingly viewed by law firms as a new vehicle for marketing and distribution of attorneys' expertise. The economic motivation of the law firms most active in development of web-based legal expert systems, such as London-based Linklaters and Sydney's Blake Dawson Waldron, and Ernst & Young, is that (1) the expert systems perform work that would not have otherwise been done by the firm, and (2) users of the expert system may become customers for more complicated, and lucrative, problems [Cohen, 1999].

Finally, funding limitations increasingly drive governmental bodies to automate delivery of legal services. There is an immense demand for routine legal information from state, federal, and local governments.[1]

Unfortunately, the AI and Law community has had little impact on the development of commercial and governmental systems for legal problem solving. For the most part, fielded systems appear to have been developed with over-the-counter software with little if any consideration of the issues that have been the primary focus of AI and Law researchers, such as models of open-texture concepts, dialectic logics, and case-based reasoning. The gulf between theory and practice in automated legal reasoning systems is potentially harmful both to commercial and government developers, who fail to profit from the years of thought and analysis embodied in the AI and Law literature, and to AI and Law researchers, whose lack the opportunity to evaluate the scalability and performance of their models in practical settings.

This paper argues that bridging the gap between the AI and Law research community and the commercial and governmental developer community will require an increased focus on rigorous task analysis, corpora of representative texts, and empirical evaluation. Section 2 sets forth a case study of precedent-based legal reasoning, arguing that failure to validate models against actual expert problem-solving behavior or representatives texts has led to a divergence between these models and actual practice in Common-Law legal jurisdictions. Section 3 discusses the requirements for AI models of legal problem solving and argues that current precedent-based reasoning systems do not constitute complete models. Section 4 argues that the absence of task analysis has led to a confusion of the actual requirements for legal problem solving with the much simpler tasks addressed by legal reasoning systems. A set of proposals for future research is set forth in Section 5.

ICAIL'03 Workshop on Evaluation of Legal Reasoning and Problem-Solving Systems, Edinburgh, UK, June 28, 2003

[1] In the United States, the NSF's Digital Government initiative is a response to this need. See http://www.digitalgovernment.org/ .

2. THE IMPORTANCE OF BEING EMPIRICAL: MISCONCEPTIONS ABOUT PRECEDENTS

This section argues that the importance of task, corpus, and empirical analysis is illustrated by the divergence of the predominant model of precedent-based legal reasoning within the AI and Law community from actual legal practice.

2.1 How Appellate Courts Use Precedents

The author of this paper was a Staff Attorney at the Colorado Court of Appeals from January 1982 through July 1985 and was a United States Supreme Court Judicial Fellow from August 2000 to July 2001. As a staff attorney, the author drafted over 180 opinions for judges of the Court of Appeals. As a Judicial Fellow, the author interacted with the Clerk and other personnel of the Supreme Court and worked with the staff of the Administative Office of U.S. Courts. This section reflects the author's experiences with appellate courts.

The Colorado Court of Appeals is typical of the lower-level appellate courts that resolve the majority of appeals and publish the majority of decisions in the United States. The Colorado Court of Appeals has initial jurisdiction, with exceptions, over appeals from the Colorado District Courts, Probate Court, and Juvenile Court. In addition, the Court of Appeals has specific appellate jurisdiction over decisions originating from a number of state administrative boards and agencies, including the Industrial Claim Appeals Office.[2] In 2002, 2,673 cases were filed in the Colorado Court of Appeals. The judges disposed of 2,463 through 291 published opinions, 1,330 unpublished opinions, 834 dismissals, and 8 transfers to the Supreme Court [Salaz, 2003].

During the author's employment at the Colorado Court of Appeals, about 75 per cent of the appeals, consisting the most routine cases, were initially directed to staff attorneys, who would read the trial record and briefs, research the issues raised in the appeal, and draft a proposed opinion with an accompanying memorandum. Occasionally, when an appeal raised an issue that had not been considered in Colorado, two alternative opinions would be drafted. The case record and proposed opinion or opinions were then assigned to a panel of judges, as were the 25 per cent of appeals that were insufficiently routine to be handled by staff attorneys.

The functioning of this court suggests several general observations that conflict with the widely-held views of researchers in precedent-based legal reasoning:

1. The outcome of disputes on questions of law is highly predictable.
2. Uncertainty arises primarily from disagreements about the characterization of facts, not from ambiguity about the precedential effect of cases.
3. Judicial decisions are almost always justified by precedent, not by statutory rules plus principles of construction.
4. Judicial decisions are not frozen dialectic.
5. Cases are not decided on the basis of teleological factors.
6. Factor-based argumentation occurs in only a small minority of cases.

Each of these observations is discussed in turn.

1. The outcome of disputes on questions of law is highly predictable. The staff-attorney system at the Colorado Court of Appeals would be unworkable unless staff attorneys could predict not only how judges will rule on a case, but also, in precise detail, the resulting decision. Experienced staff attorneys could, in fact, predict with great accuracy the outcome of the routine cases that constitute the overwhelming bulk of appeals. The least predictable appeals were cases of first impression, typically arising from new statutory provisions.

Why are so many cases with predictable outcomes appealed? In some cases, such as criminal appeals by indigents, there is no cost to the appellant from appealing. In other cases, appellants are determined to exhaust every remedy regardless of cost or probability of success. In still other cases, attorneys use litigation as a negotiating tool whose power depends more on the cost of litigation to the opposing party than on the merits of their claims. In a significant proportion of cases, the cost in attorneys' fees of performing adequate research could be a high proportion of, or even greater than, the amount in controversy. In these cases, attorneys may have only a superficial understanding of the legal issues presented by their case. Finally, empirical studies indicate that litigators systematically overestimate their liklihood of success in litigation [Loftus and Wagenaar, 1988]. Regardless of the reasons, the effect is that attorneys often bring appeals having little probability of success.

The predictability of the outcome of a case on appeal is, in general, inversely proportional to the "height" of a case in the appellate system on an axis stretching from the court with original appellate jurisdiction at the bottom to the court of highest jurisdiction at the top. In contrast to the predictability of the majority of appeals, which are brought in lower appellate courts, the minute fraction of cases that are considered by the U.S. Supreme Court are almost always highly atypical cases whose resolution is far less predictable.

2. Uncertainty in litigation arises primarily from disagreements about the proper characterization of facts, not from ambiguity about the precedential effect of cases. A key factor in the predictability of appellate cases is that appellate courts are required to accept all factual findings of the trial court for which there is any supporting evidenced, that is, unless those findings are "clearly erroneous and not supported by the record."[3] Once the facts are fixed, the outcome of a case is typically relatively clear, notwithstanding the ability of experienced attorneys to use precedents to make plausible arguments for or against the applicability of any given precedent to any given set of facts.

[2] Section 13-4-101, et seq., C.R.S. (Colorado Revised Statutes).

[3] In re Marriage of Zisch, 967 P.2d 199, 203 (Colo. App. 1998).

In practice, skillful attorneys determine the probable legal effect of alternative possible factual findings and do their utmost to persuade the trier of fact (the jury or judge, if there is no jury) to accept a characterization of the facts that leads to the legal conclusions sought by the client under the relevant legal authorities. Even on appeal, attorneys typically stretch the characterization of the facts as far as possible, consistent with the findings by the trier of fact.

Attorneys can and do argue about the meaning of cases, but this is typically a much less fruitful undertaking than arguing about the facts. Judges generally believe that they have no difficulty understanding the meaning of precedent cases. Indeed, they generally feel that they understand the law better than the attorneys practicing before them.

3. Judicial decisions are almost always by justified precedent, not by statutory rules plus principles of construction. Virtually every decision of the Colorado Court of Appeals is decided on the authority of precedent. In cases that turn on interpretation of a statutory provision, the decision cites one or more prior cases interpreting the provision. If there are no precedents in the jurisdiction interpreting the statute, precedents from other jurisdictions interpreting analogous provisions are cited. If there are no precedents anywhere, the court may discuss the legislative history or treatises by noted authorities. Principles of construction very seldom used, at least in part because canons of construction are themselves often in conflict.[4]

4. Judicial decisions are not frozen dialectic. Judges control the characterization of the issues and decide cases based on their conception of the law rather than on the particular arguments made by attorneys. In general, judicial decisions do not attempt to record all the contentions and supporting arguments of every party. To do so would be a waste of judicial resources because many arguments are meritless, redundant, or incomprehensible. At the appellate level, judicial parsimony is best served by simply rejecting the claims of the losing party, regardless of whether the court accepts the arguments of the winning party.

Far from reconstructing the dialectic history of the parties' areguments, appellate opinions typically set forth a concise description of the applicable legal doctrine as expressed in relevant precedents and explain why each contention of the appellant is wrong (in the case of affirmances) or why some contention of the appellant is correct and all contrary contentions of the appellee are wrong (in the case of reversals). The process of constructing such an opinion is very much like writing a short research monograph. The procedural setting of the case and the contentions of the parties determine the topic of the paper, but the substance of the paper is determined by legal authorities, primarily precedents, as understood by the court. It it is extremely rare for an opinion to say, in effect, that "the first party argued X because of Y, the second party argued not X because of Z, and we agree with Y rather than Z because of some general principle W".

A corollary to the observation that the analyses expressed in appellate opinions are judges' creations, rather than records of attorneys' arguments, is that oral arguments are far less important than appellant litigators commonly believe. Among U.S. federal court staff there is a saying that cases are sometimes lost, but never won, in oral arguments. The meaning of this saying is that attorneys may sometimes make fatal concessions during oral arguments, but it is exceedingly rare for an attorney to convince a judge during an oral argument of a position that the judge doesn't already hold.

At the Colorado Court of Appeals, the staff-attorney system meant that a proposed draft opinion usually already existed by the day of the oral arguments. In the U.S. Supreme Court, an oral argument provides an occasion in which all the justices are fully briefed by their clerks on the case and ready to reach a consensus among themselves as to the decision. Questions directed to the attorneys during arguments are often covert discussions among the justices, e.g., one justice will interrupt an attorney to say, "I would have thought that you would argue X", to which another justice will ask, "Yes, but then how would you respond to Y?", and so on, with the hapless attorney acting simply a surrogate for contending justices.

5. Cases are seldom decided on the basis of teleological factors, such as general principles or policies. The importance of telelogical factors in legal discourse was emphasized in an influential paper by Berman and Hafner [Berman and Hafner, 1993]. Berman and Hafner (1993) made the accurate observation that the arguments produced by automated legal reasoning systems, such as GREBE, differed from those of an experience attorney in that they focused on factual differences between cases without relating those factual differences to the purposes underlying the applicable legal doctrine. Berman and Hafner argued that the output of systems should be framed in terms of these purposes. However, this paper has subsequently been interpreted as making the much stronger assertion, not just that legal discourse should relate factual similarities and differences between cases to the purposes underlying legal doctrines, but that precedent-based decisions should actually be made on the basis of teleological factors. Although teleological factors are sometimes important in cases at the top end of the hierarchy of appellate courts, in routine cases they play little role in judicial decision making and are almost never mentioned in any but the most cursory fashion.

There are two main reasons that teleology plays such a minor role in decision making (though it may appear in appellate discourse and argumentation). First, there is often no consensus as to the abstract principles justifying decisions.[5] For example, a given criminal sentence may be justified by

[4]Posner describes canons of construction as

> Cautionary rather than directive, often pulling in opposite direcitons like their counterparts, the maxims of ordinary life ("haste makes waste," but "he who hesitates is lost"), the canons are the collective folk wisdom of statutory interpretation and they no more enable difficult questions of interpretation to be answered than the maxims of everyday life enable the difficult problems of everyday living to be solved. [Posner, 1990] pg. 280.

[5]This argument, and the sentencing and strict liability examples below, appear in [Sunstein, 1996].

retribution, rehabilitation, deterrence, prevention, or some mixture of these policies. As a practical matter, it is usually impossible for a society to reach a consensus about the relative importance of these factors. However, there is usually a consensus that similar crimes should have similar punishments. Thus, a sentence may be justified by precedent even if the the underlying theory of sentencing is in dispute. Similarly, strict financial liability to distributors of contaminated food is favored by some on the grounds of economic efficiency, by others to achieve distributive goals, and by still others to promote a basic human right to safe food. Despite their disagreements, advocates of these conflicting positions may all agree that distributors of contaminated food should be held strictly liable for any harm resulting from food contamination. In general, there can be agreement about specific cases even when, as is often the case in a democratic society, there is disagreement about the rationale for those specific cases. More abstract goals simply make balancing more intractable. For example, can one really imagine justifying a sentence based on a general judgment about the relative importance of mercy and justice?

Second, even if there were a consensus about the relative importance of competing high-level goals, it is often impossible to determine what decision best promotes these goals. For example, it would probably be unfeasible to make a choice between parole and prison for a given criminal based on a trade-off between the costs of incarceration and the criminal's family's lost income against the lost economic productive resulting from by crime by paroled criminals. A calculation of such factors is simply beyond the capabilities of any court.

However, regardless of the reasons, as a matter of fact most judicial decisions simply do not contain language justifying the outcome on the basis of high-level societal goals.

6. Factor-based argumentation occurs in only a small minority of cases. Factor-based argumentation is an important form of argument in those areas of law for which there are no well-defined rules, but in which courts have identified a collection of relevant factors. Areas of law having this characteristic include certain areas of tax law, such as the "home office deduction", the U.S. Emergency Petroleum Allocation Act of 1973, and intellectual property law.[6] In these *factor-governed* domains, arguments are based on the balance among competing factors as reflected in the "totality of the circumstances." Precedents provide reference points at which the result of balancing particular sets of factors is known.

A substantial number of cases involve factor-based argumentation. Cass Sunstein, in *Legal Reasoning and Political Conflict* [Sunstein, 1996], estimates that the number of U.S. federal cases between 1944 and 1994 involving factor-based argumentation to be 11,155, based on a Lexis search for the phrase "totality of the circumstances."

However, by the mid-1990's the U.S. District Court was deciding over 250,000 cases per year and the U.S. Courts of Appeal were deciding over 50,000 [Meachem, 1996]. Thus, the number of cases involving the "totality of the circumstances" in the 50-year interval ending in 1994 was apparently only about 1/30 of the number of federal cases decided on the last year of that interval. Evidentally, only a very small proportion of U.S. federal cases are decided on the basis of the "totality of the circumstances." This is consistent with my observation that in the Colorado Court of Appeals cases involving balancing competing factors were very uncommon. While factor-based argumentation is widespread and important, it nevertheless occurs only in a minority of cases.

[6] However, it is arguable that in jurisdictions in which the Uniform Trade Secrets Act has been adopted, intellectual property law is relatively rule-governed [McCarty, 1997].

If these contentions are correct, there is striking inconsistency between the assumptions underlying many recent formal models of precedent and actual practice. Arguably, this inconsistency has arisen because of the failure of researchers to validate their models against actual expert problem-solving behavior and non-trivial numbers of representatives texts.

3. AI MODELS OF LEGAL PROBLEM SOLVING

A theoretical model has been described as consisting of:

> entities (objects, materials, mechanism, systems, structures) belonging to a relatively unproblematic, more familiar, or better-organized secondary domain. ...
>
> Inferences from the assumptions made in the secondary field are translated by means of ... rules of correlation ... [into] the primary domain. ...
>
> [T]he identity of structure ... permits assertions made about the secondary domain to yield insight into the original field of interest. [Black, 1962] at pg. 230.

An artificial intelligence model consists of a representation of stages in a problem-solving process in the form of configurations of symbols, together with symbolic operators that transform one configuration into another. "Rules of correlation" translate from a real-world problem into a symbolic representation of the model's start configuration, and from the model's goal configuration into a solution to the real-world problem. AI models are typically evaluated by such factors as their effectiveness at deriving solution representations from problem representations, the computational tractability of the algorithms employed in the derivation, and the acquisition cost of the knowledge required for the algorithms.

Under this view of what constitutes an AI model, the majority of AI and Law formalisms and systems are not complete AI models because they don't fully specify the relationship between the symbolic representations and actual legal discourse. For example, systems such as GREBE [Branting, 2000] and HYPO [Ashley, 1990] generate textual output with some resemblance to legal memoranda, but neither specifies the relationship between the input of the system and the discourse that might be produced by a user in describing a problem under any realistic problem-solving scenarios. Similarly, the sophisticated legal concept models proposed by McCarty [McCarty and Sridharan, 1982, McCarty, 1995] are not accompanied by any explanation of how these models arise from, and give rise to, the language actually used by attorneys when solving problems. Most logical models of legal argumentation make no attempt to model legal discourse or problem solving *per se*, and indeed have been criticized on precisely those grounds [McCarty, 1997].

In fairness, none of these projects explicitly purported to provide a complete model of legal problem solving. For example, logical models of legal argumentation are typically intended to be normative models to be evaluated on the basis of formal criteria, rather than empirical models of legal problem-solving behavior. However, the failure to specify exactly what is being modeled, that is, the "rules of correlation" between the model and the thing being modeled, gives rise to confusion about precisely what the research contribution of a research project consists of and precisely what is left undone.

To illustrate this point, I will turn to the example of HYPO [Rissland et al., 1984, Ashley, 1990] and its progeny, CATO [Ashley and Aleven, 1997, Aleven and Ashley, 1997], and CABARET [Skalak and Rissland, 1992, Rissland and Skalak, 1991] since these are the best-known projects in automated precedent-based legal reasoning. HYPO models case-based argumentation in factor-governed domains. In HYPO's representation, there is a direct correspondence between the legal factors identified by courts as determinative of the outcome of cases and the symbolic dimensions or factors in terms of which cases are represented. HYPO's model of argumentation is based on pairwise comparisons of the collections of symbolic factors favoring each party in different cases.

There is a close correspondence between HYPO's model of argumentation and discourse that could actually occur in cases in factor-governed domains. Thus, HYPO constitutes a plausible, testable model of actual human case-based argumentation in factor-governed domains, such as those in which HYPO, CATO, and CABARET were implemented. While no systematic evaluation of HYPO as a model of actual case-based argumentation has, to the best of my knowledge, been performed, the effectiveness of a tutorial system having HYPO as its model of expertise has been evaluated, with some tentative, but promising, results.

Outside of the developers of HYPO, CATO, and CABARET, however, the HYPO model has often been assumed to be not just a plausible, but largely untested, model of argumentation in factor-based domains, but instead to be the definitive model of precedent-based legal argumentation in all domains, factor-based or not. Numerous ICAIL papers fail even to mention the existence of alternative models of precedent or to provide any justification for applying a formal model developed for factor-based domains to non-factor based domains. For example, one recent ICAIL paper stated without qualification that "The most useful representation of cases is as factors " [Bench-Capon and Sartor, 2001].[7]

The implicit hypothesis that factor-based argumentation is superior to alternative computational models of precedent-based argumentation (such as any of the numerous computational models of analogy, e.g., [Meldman, 1975], [Haraguchi, 1993], [Yoshino et al., 1993], [McCarty and Sridharan, 1982], [Indurkhya, 1997], or [Branting, 2000]) even in non-factor-based domains, has never been rigorously tested, much less established. In legal domains in which courts have not articulated factors, the correspondence between the computational model and legal discourse is much less direct than in factor-governed domains. Under these circumstance, the symbolic factors used in the representation of cases are knowledge-engineering artifacts, rather than entities recognized in primary legal authorities.

Several papers have addressed the applicability of the HYPO model to non-factor-governed domains. *An Attempted Dimensional Analysis of the Law Governing Government Appeals in Criminal Cases* [Mendelson, 1989] found that "dimensional analysis does not succeed in the domain of 18 U.S.C. 3731 jurisprudence," primarily because the "unsettled, fact-sensitive, and conflicting" nature of the domain meant that "dimensions cannot fully capture the true emphasis of the decisions." Similarly, [Branting, 1994] and [Branting, 2000] argued that in the domain of Torts, factors and cases are insufficient in themselves to distinguish alternative theories of decision. Alternative theories of decision require, in effect, separate case representations in terms of alternative sets of factors. Thus, the alternative theories of decision of a case cannot be simultaneously represented.

Some legal scholars have treated factor-based argumentation as wholly distinct from reasoning by analogy, arguing that the former is applicable when courts have articulated a fixed set of relevant factors but have not described how the factors interact, whereas analogy occurs when decisions depend on "reflective equilibrium" [Sunstein, 1996]. Under this view, factor-based argumentation and reasoning by analogy are appropriate in entirely different legal domains.

In summary, HYPO's model of case-based argumentation is a plausible model for factor-governed legal domains because in such domains there is a clear correspondance between the model's symbolic elements and entities recognized by courts as relevant to the outcome of cases. Little research has addressed the issue whether this model corresponds closely to legal discourse in non-factor governed domains, but there is

[7] It is instructive to contrast this statement with the widespread view among legal scholars as to the importance of case facts:

- "[I]t is the facts, and not the general rules of law found in precedents ... which serve as the foundation of the decisional process." [Cueto-Rua, 1981] at pg. 56.
- "Judgments must be read in light of the facts of the case in which they are delivered" [Cross, 1991].
- "It is clear that the most important part of a representation for cases is the representation of the facts and the outcome." [Gardner, 1987] at pg. 47. Gardner quotes with approval Corbin's admonition that cases should be studied "not so much for their doctrinal statements as for ... their operative facts" and Gilmore's statement that one should "never quote general language from an opinion, divorced from the factual context in which the language was delivered."

clearly some evidence that it does not.[8] Resolving the issue of the extent to which HYPO, or any other formalization of legal reasoning, models expert problem-solving behavior and discourse requires considering the specific tasks that experts perform and the actual discourse that they receive and produce.

4. TASK ANALYSIS IN LEGAL PROBLEM SOLVING

The history of AI and Law has been punctuated by unresolved clashes of intuition about computational models of legal reasoning[9] A central reason that these disputes remain unresolved is the absence of rigorous analyses of the tasks that arise in legal problem solving. In the absence of empirical evidence of the actual behavior of legal problem solvers, there are few grounds for resolving clashing intuitions.

In this respect, AI and Law contrasts with AI and Medicine (AIM). In the early years of research in AIM there were numerous rigorous studies of expert behavior.[10] These studies played a critical role in directing AIM research. Throughout its history, knowledge acquisition has been recognized as key to the success of AIM systems [Tu et al., 1995]. Lamentably, there have been comparatively few empirical studies of decision making by lawyers (although jury behavior has been studied extensively, e.g., [Penrod et al., 1990]).

A variety of legal tasks can be distinguished, including advocacy, legal planning, adjudication, counseling, and drafting legal documents. The requirements of each of these tasks depends critically upon the characteristics and goals of the person for whom the task is performed. Potential consumers of legal expertise include at least the following:

- Litigants - those affected by legal decisions, including:
  - individuals and groups of individuals
  - governmental and commerical institutions
- Attorneys - agents representing litigants
- Judges - decision-makers for legal questions
- Juries - decision-makers for factual questions
- Clerical staff - those who manage case-specific documents and information
- Legislators and regulatory rule-makers
- Scholars, such as legal philosophers and law professors

As an example of the extent to which the nature of a legal task is shaped by the needs of the user, consider the following two legal queries, the first a paraphrase of an actual problem posed to the author when he was a practicing attorney, the second the same problem as it might be formulated by an attorney.

> **Query 1.** At the place where I work, our bookkeeper didn't give me my paycheck last month. Instead, she signed my name on it, cashed it, and left town. I don't know where she went. I asked my boss to give me a new check for my salary, but he said that he had already paid me once and that he didn't have to pay me again. He says that if anyone owes me the money, it is the bookkeeper. Is he right that he doesn't owe me my wages anymore?

> **Query 2.** Under Article 3 of the Uniform Commercial Code, is a payor's obligation to a payee discharged by a negotiable instrument if the negotiable instrument is paid to a third party over a forged signature?

The task of answering Query 1 is radically different from (although it subsumes) the task of answering Query 2. Indeed, most of the expertise needed to answer Query 1 is required for reformulating the problem as Query 2. Once expressed in the form of Query 2, the problem is relatively trivial.

Research projects that purport to answer queries of the first type using computational methods applicable only to queries of the second type are indulging in a rhetorical slight of hand in which almost all of the expertise has been squeezed into the representation of problem, leaving only the simplest final steps for the computer to perform.

Most precedent-based systems are guilty to some extend of this slight of hand. Continuing with the example of HYPO (but without suggesting that it is unusual in this respect) consider the analog of a type-2 query in HYPO. Suppose that attorneys for plaintiff and defendant are litigating in a factor-governed domain and agree on the applicable factors. Each then cites cases supporting the client's position, employing the rhetorical moves well-described in the HYPO literature, such as distinguishing, trumping, etc. When one party establishes a monopoly on non-trumped points, the judge rules in favor of that party.

The world would, perhaps, be a better place if attorneys were so well-behaved. In reality, however, the attorneys in this scenario would make a quick calculation of what points would be non-trumpable under various sets of factors arguably justifable under the current factual situation and the applicable cases. Based on this calculation, each would aggressively argue about what factors were actually satisfied. For example, the sorts of arguments attorneys might make concerning whether the factor "There is data that plaintiff made disclosures to outsiders" [Ashley, 1990] is present might include how much and what type of information must be revealed to constitute a disclosure, whether unintentionally revealing information constitute a disclosure and if not what constitutes "intentional," whether reasonable care was taken to prevent a disclosure, what constitutes reasonable

[8] It is, of course, possible to construct factor-based arguments in non-factor-governed domains. See, e.g., [Rissland, 1989]. However, this doesn't in itself establish that attorneys in fact make such arguments nor that judges (the final arbiters of persuasiveness in the judicial system) find such arguments persuasive.

[9] See, e.g., [McCarty, 1997].

[10] See http://smi-web.standord.edu/pubs/yearindex.html.

care in the context of the case, whether the category "outsiders" include technical staff, contractors, members of the board of directors, or employees of subsidiaries, etc. The number of arguments that a skillful attorney could make about the presence or absence of this or any other factor is essentially unlimited.

The fact that attorneys would need to be able to determine the consequences of establishing various factors confirms that the subject matter of CATO is a real legal skill that law students can and should learn. However, this skill is less important than lawyers' ability to characterize a given factual situation in whatever manner is most likely to achieve a client's goals under the applicable legal authorities. In other words, the most important task is answering type-1 queries.[11] Answering type-2 queries is an important component of this task, but it is a serious error to confuse type-2 query answering with the much more difficult type-1 task.

In summary, the failure of AI and Law research projects to explicate the actual tasks that they address has led to a conflation of relatively simple type-2 queries that most systems actually answer with the much harder type-1 queries that many users require.

5. RECOMMENDATIONS

The AI and Law community can play an important role in the ongoing transformation of the legal services from a labor-intensive, manual activity to a highly automated industry. However, this role will require us to change our focus away from desk-top demonstrations and towards empirical analysis of real-world problem solving and toward scalable, validatable systems.

Here are some suggestions on steps we can take to help achieve these goals:

- Focus on analysis of economically significant tasks, then develop algorithms and representations that address those tasks in the context of real-world problems.
- Perform rigorous protocol and corpus analyses to find out what practical problem solving requires.
- Develop techniques usable by developers outsider of our immediate community and promote them to these developers.
- Expand beyond a handful of examples by developing corpora, performing statistical analysis, and building systems that are used by someone other than the developers. Follow the example of IAAI in providing formal recognition to projects that have achieved independence from their original developers.
- Perform ablation studies to separate out the contributions of the various components of systems, to avoid conflating the effects of separate components of the system, e.g., representation with algorithms.
- Don't ignore knowledge acquisition and problem elicitation (i.e., client interviewing). For example, if CBR requires more powerful KA techniques than are available right now, either focus on developing such techniques or shift to other, more tractable, tasks.
- The relevance of AI and Law work to actual legal discourse and problem solving should always be made apparent. A successful AI model should always permit one to solve a problem that could not be otherwise be solved, perform a task that could not otherwise be performed, or answer a question that could not be answered, without the model.
- The acid test AI models should be persuasiveness to experts who are familiar with the domain being modeled but who have no personal, institutional, or intellectual stake in the model.

The assertions in this paper concerning the use of precedent will no doubt be counterintuitive to many readers because they contradict assumptions widespread in the AI and Law community. I urge those who doubt these assertions to adduce whatever evidence is required to prove them wrong. But I urge even the most skeptical reader not to simply assume without evidence that the assertions are false. Base your conclusions on actual legal discourse as set forth in transcripts and primary legal authorities.

Only when our conclusions are rooted in actual legal practice are we likely to make a significant contribution to the legal community and the citizen-consumers of legal services.

[11]The importance of reasoning with specific facts, rather than simplified abstractions, was stressed above in item 2, Section 2.1, (uncertainty in litigation arises primarily from disagreements about the proper characterization of facts, not from ambiguity about the precedential effect of cases) and in footnote 7, which quotes a series of legal scholars who stress the centrality of the specific facts of precedents.

6. REFERENCES

Aleven, V. and Ashley, K. (1997). Evaluating a learning environment for case-based argumentation skills. In *Proceedings of the Sixth International Conference on Artificial Intelligence and Law*, pages 170–179, University of Melbourne, Melbourne, Australia. ACM Press.

Ashley, K. (1990). *Modelling Legal Argument: Reasoning with Cases and Hypotheticals.* MIT Press, Cambridge, Massachusetts.

Ashley, K. and Aleven, V. (1997). Reasoning symbolically about partially matched cases. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence*, pages 335–341, San Francisco. Morgan Kauffmann.

Bench-Capon, T. and Sartor, G. (2001). Theory based explanation of case law domains. In *Proceedings of the Eighth International Conference on Artificial Intelligence and Law*, pages 12–21, St. Louis, Missouri.

Berman, D. and Hafner, C. (1989). The potential of artificial intelligence to help solve the crisis in our legal system. *Communications of the ACM*, 32:928–938.

Berman, D. and Hafner, C. (1993). Representing teleological structure in case-based legal reasoning: The missing link. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Law*, pages 50–59, Amsterdam, The Netherlands. ACM Press.

Black, M. (1962). *Models and Metaphors.* Cornell University Press, Ithaca, NY.

Branting, K. (1994). A computational model of *ratio decidendi*. *Artificial Intelligence and Law*, 2(1):1–31.

Branting, K. (2000). *Reasoning with Rules and Precedents: A Computational Model of Legal Analysis.* Kluwer Academic Publishers, Dordrect/Boston/London.

Cohen, A. (1999). Legal advice without the lawyers. *New York Law Journal.*

Cross, R. (1991). *Precedent in English Law.* Oxford University Press, Oxford, fourth edition.

Cueto-Rua, J. C. (1981). *Judicial Methods of Interpretation of the Law.* The Publications Institute, Paul M. Hebert Law Center, Louisiana State University.

Gardner, A. (1987). *An Artificial Intelligence Approach to Legal Reasoning.* Bradford Books/MIT Press, Cambridge, MA.

Haraguchi, M. (1993). A reasoning system for legal analogy. *Machine Intelligence*, 14:329–352.

Indurkhya, B. (1997). On modeling creativity in legal reasoning. In *Proceedings of the Sixth International Conference on Artificial Intelligence and Law*, pages 180–189, University of Melbourne, Melbourne, Australia. ACM Press.

Loftus, E. and Wagenaar, W. (1988). Lawyers' predictions of success. *Jurimetrics*, pages 437–453.

McCarty, L. T. (1995). An implementation of Eisner v. Macomber. In *Fifth International Conference on Artificial Intelligence and Law*, pages 276–286, College Park, MD. ACM Press.

McCarty, L. T. (1997). Some arguments about legal arguments. In *Proceedings of the Sixth International Conference on Artificial Intelligence and Law (ICAIL-97)*, pages 215–224, University of Melbourne, Melbourne, Australia. ACM Press.

McCarty, L. T. and Sridharan, N. S. (1982). A computational theory of legal argument. Technical Report LRP-TR-13, Laboratory for Computer Science Research, Rutgers University.

Meachem, L. (1996). Judicial business of the united states courts. Technical report, Administrative Office of the United States Courts.

Meldman, J. A. (1975). *A Preliminary Study in Computer-Aided Legal Analysis.* PhD thesis, Massachusetts Institute of Technology.

Mendelson, S. (1989). An attempted dimensional analysis of the law governing government appeals in criminal cases. In *Proceedings of the Second International Conference on Artificial Intelligence & Law (ICAIL-89)*, pages 128–137, Vancouver, BC.

Morrison, R. (1989). Market realities of rule-based softrware for lawyers. In *Proceedings of the Second International Conference on Artificial Intelligence and Law*, pages 33–36, Vancouver, B.C.

Penrod, R., Hastie, S., Pennington, N., and Penrod, S. (1990). *Inside the Jury.* Harvard University Press.

Posner, R. (1990). *The Problems of Jurisprudence.* Harvard University Press.

Rissland, E. (1989). Dimension-based analysis of hypotheticals from supreme court oral argument. In *Second International Conference on Artificial Intelligence and Law*, pages 111–120, Vancouver, B.C. ACM Press.

Rissland, E. and Skalak, D. (1991). Cabaret: Rule interpretation in a hybrid architecture. *International Journal of Man-Machine Studies*, 34(6):839–887.

Rissland, E., Valcarce, E., and Ashley, K. (1984). Explaining and arguing with examples. In *Proceedings Fourth National Conference on Artificial Intelligence (AAAI-84)*, pages 288–294, Austin, TX.

Salaz, K. (2003). Personal communication. Office of Colorado State Court Administrator.

Skalak, D. and Rissland, E. (1992). Arguments and cases: An inevitable intertwining. *Law and Artificial Intelligence*, 1(1).

Sunstein, C. (1996). *Legal Reasoning and Political Conflict.* Oxford University Press.

Susskind, R. (1998). *The Future of Law.* Oxford University Press, Oxford, second edition.

Tu, S., Eriksson, H., Gennari, J., Shahar, Y., and Musen, M. (1995). Ontology-based configuration of problem-solving methods and generation of knowledge-acquisition tools: application of protege-ii to protocol-based decision support. *Artificial Intelligence in Medicine*, 7(3):257–289.

Yoshino, H., Haraguchi, M., Sakurai, S., and Kagayama, S. (1993). Towards a legal analogical reasoning system: Knowledge representation and reasoning methods. In *Fourth International Conference on Artificial Intelligence and Law*, pages 110–116, Amsterdam, The Netherlands. ACM Press.