Noname manuscript No.
(will be inserted by the editor)

A History of AI and Law in 50 papers:

25 Years of the International Conference on AI and Law

Trevor Bench-Capon · Michał Araszkiewicz · Kevin Ashley · Katie Atkinson · Floris Bex · Filipe Borges · Daniele Bourcier · Paul Bourgine · Jack G. Conrad · Enrico Francesconi · Thomas F. Gordon · Guido Governatori · Jochen L. Leidner · David D. Lewis · Ronald P. Loui · L. Thorne McCarty · Henry Prakken · Frank Schilder · Erich Schweighofer · Paul Thompson · Alex Tyrrell · Bart Verheij · Douglas N. Walton · Adam Z. Wyner

the date of receipt and acceptance should be inserted later

Trevor Bench-Capon · Katie Atkinson · Adam Z. Wyner
Department of Computer Science, University of Liverpool, UK.
E-mail: tbc@liverpool.ac.uk

Michał Araszkiewicz
Department of Legal Theory, Jagiellonian University, Cracow, Poland

Kevin Ashley
University of Pittsburgh, Pittsburgh, PA, USA

Floris Bex
University of Dundee, Dundee, UK and University of Groningen, The Netherlands

Filipe Borges
Legal Information Systems, Paris, France

Daniele Bourcier
Law and complex systems, CNRS, Paris, France

Paul Bourgine
Reseau national des systemes complexes, Paris, France

Jack G. Conrad · Jochen L. Leidner
Thomson Reuters Global Resources, Catalyst Lab, Baar, Switzerland

Enrico Francesconi
Institute of Legal Theory and Techniques, Italian National Research Council (ITTIG-CNR), Florence, Italy

Thomas F. Gordon
Fraunhofer FOKUS, Berlin, Germany

Guido Governatori
NICTA, Queensland Research Laboratory, Australia

David D. Lewis
David D. Lewis Consulting, Chicago, USA

Ronald P. Loui
Department of Computer Science, University of Illinois-Springfield, USA

L. Thorne McCarty

Contents

Department of Computer Science, Rutgers University, New Brunswick, New Jersey, USA

Henry Prakken
Utrecht University and University of Groningen, The Netherlands

Frank Schilder · Alex Tyrrell
Thomson Reuters Corporate Research and Development, USA

Erich Schweighofer
Centre for Computers and Law, Faculty of Law, University of Vienna, Austria

Paul Thompson
Dartmouth College, Hanover, NH, USA

Bart Verheij
Department of Artificial Intelligence, University of Groningen, The Netherlands

Douglas N. Walton
Centre for Research in Reasoning, Argumentation and Rhetoric, University of Windsor, Windsor, Ontario, Canada

Abstract We provide a retrospective of twenty-five years of the International Conference on AI and Law, which was first held in 1987. Fifty papers have been selected from the thirteen conferences and each of them is described in a short subsection individually written by one of the twenty-four authors. These subsections attempt to place the paper discussed in the context of the development of AI and Law, while often offering some personal reactions and reflections. As a whole, the subsections build into a history of the last quarter century of the field, and provide some insights into where it has come from, where it is now, and where it might go.

1 Introduction

The first International Conference on AI and Law was held in Boston in May 1987. Since then it has been held every odd numbered year, and so 2011 saw the thirteenth edition of the conference and 2012 marks its twenty-fifth anniversary. Although there had been some earlier work on AI and Law, the first ICAIL can be seen as birth of an AI and Law *community*. Other signs of an emerging community followed: in Europe the Jurix conferences have been held annually since 1988, the International Association for AI and Law had its inaugural meeting at the Third ICAIL in 1991, and this journal was first published in 1992. Thus twenty five years seems like a good time to take stock of the field, and a retrospective comprising papers published at ICAIL is a good way to do this. Here we have brought together a group of people from the AI and Law Community and asked them to write a short appreciation of papers that have meant something to them. This work therefore largely consists of a discussion of fifty papers that first appeared at ICAIL, but in so doing it provides something of a history of the field, and an opportunity to reflect on progress made and lessons learned.

ICAIL has played an important part in the development of AI and Law. Many of the key ideas of the field were introduced at an ICAIL conference and further developed over successive conferences. Trends have been started, and topics have waxed and waned. A number of factors have combined to give ICAIL this role.

- It takes place every other year, rather than every year. This means that it remains somehow special, but more importantly it allows for development of ideas between conferences. Whereas with an annual conference the call for papers for the next edition is often distributed at the conference, and many people are already working on what they will present at the next conference, alternate years gives an opportunity to reflect on what was presented at a conference before starting to write for the next one, and this allows scope for cross-fertilisation. A PhD student who presents some initial findings at one conference will be on the verge of submitting completed research by the next. Annual conferences can blur into one another; ICAIL has managed to retain a sense of development between conferences.
- It is the only truly international conference in AI and Law. There are of course local and regional gatherings of AI and Law people, but ICAIL is where the worldwide community gathers. This reinforces the sense that ICAIL is special, but importantly allows for ideas to be communicated to the whole community at a single time, and to get some wide ranging feedback on the ideas. This in turn means that people want to present their best work at ICAIL: it provides the big stage for dissemination.
- ICAIL has always retained only plenary sessions. This is important in that it expresses a philosophy that all the various applications and approaches can contribute to one another. Too many conferences rely on highly parallel sessions, which have a tendency to fragment and factionalise the field. As will emerge in this paper, one of the notable developments in AI and Law has been the growing recognition of the contribution that logic-based and case-based approaches can make to one another. This recognition has been fostered by ICAIL and its plenary structure: had logic and cases been compartmentalised into different parallel sessions, this mutual awareness and understanding might never have developed.
- ICAIL is a good size: big enough that the worldwide field is properly represented, but small enough that it is possible to meet everyone one wishes to.

It is always good to take stock of a field: as George Santayana told us, those who do not remember the past are condemned to repeat it. Moreover a retrospective view enables us to consider the work of the past from the perspective of the present: we now know how the ideas will develop and what they will lead to; which ideas will flourish and which will not. Even more importantly we can often understand ideas that we had only a dim grasp of at the time. Sometimes too our understanding of a paper changes: just as legal cases may need to be reinterpreted in the light of subsequent decisions (e.g. [158]), so too research papers may need some reinterpretation in the light of what comes later. Moreover, technology advances apace in Computer Science: some ideas which could not be realised in 1987 can be implemented easily now: others may need to be rethought in the light of the way the technology has developed. The emergence of the World Wide Web in particular has had an enormous impact on legal informatics, both what is feasible and what is desirable.

This work therefore should be visited like a retrospective of an artist: it is interesting to see what was done and how that was developed, but the point is to understand the past so as to better anticipate the future.

1.1 Structure. *Trevor Bench-Capon*

The remainder of this work comprises a series of sections, each devoted to a particular conference. After a brief note on the conference itself (written by myself), giving some facts about the conference and recalling some personal highlights, the subsections describe particular papers presented at that conference. Each subsection was written by a particular person, named in the subsection title, and reflects that person's subjective view. The idea is to gather a diversity of voices and to reflect a range of opinions, interests, backgrounds and preferences. The authors all chose the papers that they write about, as being of some particular significance to them. The final section looks ahead to ICAIL 2013 in Rome, and is written by the Programme Chair of that Conference.

There are twenty-four different authors for the fifty papers, representing a dozen different countries. Some of the authors have attended every ICAIL and some only one. Some are beginning their career and some are coming to the end of theirs. Some work in academia, others in a commercial environment. By bringing these different opinions together in this way we hope to capture something of the whole range of AI and Law as it has developed since 1987, and the different opinions that can be found in the community.

2 Boston 1987

The very first ICAIL was held from May 27th to 29th 1987, in Boston, Massachusetts, at Northeastern University. Carole Hafner was Conference Chair and Thorne McCarty was Programme Chair. Strikingly there were only five programme committee members, Donald Berman, Michael Dyer, Anne Gardner, Edwina Rissland and, the only European, Marek Sergot. John McCarthy gave the Banquet Address (actually his standard seminar on circumscription). Already at this conference some key themes were evident. A number of papers related to conceptual retrieval, and three of the four papers chosen for this section relate to this topic, perhaps reflecting the then greater maturity of work in this area since it had developed from Information Retrieval, which had been an early application of computers to Law. One of these introduced artificial neural networks as a way to handle incompleteness and apparent inconsistency which are central features of the legal domain. The fourth paper in this section concerns the use of a formalisation of legislation to build an expert system. But case-based systems were also represented at the conference, notably [227] which introduced HYPO - we would hear much more of this program and its descendants. The roots of rule-based argumentation can be found in a paper such as [99], which stressed the need to acknowledge the nonmonotonic behaviour introduced by the use of exceptions to general provisions in much legal drafting. There was also a panel session which presented different views on formal and computational approaches to modelling legal reasoning. In this panel the different approaches, rule-based versus case-based, and classical logic versus nonmonotonic rule systems, were set up as competitors. That was the view back in 1987: one of the achievements of ICAIL has been that these views are now recognised as complimentary rather than opposed.

2.1 Richard K. Belew: A Connectionist Approach to Conceptual Information Retrieval [26]. *Commentary by Filipe Borges, Daniele Bourcier and Paul Bourgine*

In AI, the eighties were mainly concerned with symbolic representation. Belew's paper [26] was one the first in AI and Law to introduce the connectionist approach, often termed the sub-symbolic approach. Similar to the functioning of the brain, sub-symbolic knowledge emerges in a "serendipidous" (*sic* Belew, in his conclusion) way from the neural network. Another important original contribution of this paper was to argue for the relevance of this type of representation for the open textured concepts typical of the Law (e.g. Hart [129]).

Briefly speaking, whereas traditionally retrieval systems had been based on simple words and indexes, this paper advocated a new method for capturing concepts for retrieval. However, the particular legal application was not the crucial point, and the interest in neural networks has resulted in various other interesting applications exploiting their ability to learn information, and to determine similarity between cases. Indeed the most creative aspect of this report concerned the new "paradigm" of hybrid knowledge, integrating symbolic and non symbolic representation. Representation of the decision making process became one of the main applications of artificial neural networks, inherited from Peirce's work on abduction, deduction and induction [197].

Given the many failures of the various attempts to model abstract legal phenomena by means of symbolic approaches, the connectionist approach introduced in Belew's paper and demonstrated in subsequent works offers a particularly rich analysis framework. Like classic expert systems, the learning base of connectionist expert systems can be built from complete and consistent set of rules, but connectionist expert systems can also be based on incomplete sets of rules or cases, structured with fuzzy descriptive criteria, which contain inconsistencies and even contradictions.

In other words, these connectionist expert systems are resilient in the face of the incompleteness and inconsistency that characterizes the legal field. Their internal structure allows them to somehow absorb these inconsistencies and to construct a representation of the phenomena which can be very instructive. It is thus possible to observe in these models the emergence of legal categories known or new, or to discover optimized methods of representing law or solving legal problems. It is also possible to use them for analysis of legal corpora or legal decision making, to reveal the principles, contradictions or gaps, or to detect evolving trends and so use these models to explore the evolution of legal cases.

Finally, and contrary to some subsequent conclusions, further work will demonstrate that it is possible to explain the reasoning used by these models, whether sub-symbolic or hybrid. Thus, these justifications, which can be given even in a fuzzy environment, provide a framework for analyzing the decision-making process.

Whatever the method for constructing the connectionist expert system, it has been proved that it produces - given the qualified facts - Bayesian decisions, when the number of cases is sufficiently large. The Bayesian is the best nonlinear predictor given the data and it is not dependent on the method for constructing it. The complete Bayesian partitions the space of qualified facts into basins of attraction for the different qualitative decisions and, inside each basin produces the qualitative sub-Bayesian of quantitative decision. Each basin of attraction can be represented automatically as a set of expert rules.

The basins of attraction can be also represented equivalently and automatically as a decision tree. In the second case the nodes of the decision tree make distinctions that are very relevant hierarchically for justifying a decision. The main problem of neural nets as it has been stressed resides in the statistical representativity of data.

Neural networks can offer more than a metaphor for law as a parallel process: they are not just a pragmatic way to extract rules. Future research that creates such dynamic reasoning systems will test implications of various theories on the evolution of law as complex adaptive systems.

The whole learning process will always contain uncertainties if the number of cases is low and if the description of the qualified facts are poor. In the era of "Big data", the best way to produce a connectionist expert system is to design and build a social intelligent ICT System, in the manner of Wikipedia. Such an expert system will progressively assimilate the expertise of a large community of involved (legal) experts by active learning "Juripedia".

With sophisticated interfaces it will become a valuable help for all people involved in the legal process, lawyers and non-lawyers alike.

2.2 Carole D. Hafner. Conceptual Organization of Case Law Knowledge Bases [114]. *Commentary by Adam Z. Wyner*

The legal case base contains thousands of cases, all represented as text. In the last forty years or so, the corpora of cases have been digitally represented, enabling digital searches. With new technologies come new opportunities. For the case base, this was the opportunity to search quickly and efficiently through the cases for particular information, facilitating the identification of relevant cases for academic study or legal argument. One prominent approach since the 1960s was simply Boolean searches, using “and”, “or”, and “not”, for matching strings; companies such as LexisNexis and Westlaw provided legal firms with databases and search facilities, showing it to be a commercially viable approach. To the current day, string search continues to be the primary search approach used by web-based search engines, e.g. Google, albeit vastly improved with the use of pre-indexation of web pages on the Internet.

However, string searches are limited in a range of respects. In [114], several of these limitations are highlighted, and an alternative approach is proposed - concept indexation and concept search. Among the limitations of string search are that it requires some expertise to know what search terms to use, that it is not feasible to know all the relevant combinations of terms, and that there is an over-estimation of the quality of the results. A better approach would be to annotate the cases with conceptual cover terms, which are semantically meaningful generalisations over the textual particulars and variant expressions - textual differences would be homogenised, and relationships between textual elements could more easily be identified. Then searches would be with respect to fewer terms or related terms, providing a better, more accurate rate of returned cases. However, to annotate the cases requires that the textual content be annotated. Prior to the annotation task, we must provide some “map” of the relevant legal concepts and issues and their relationships to the legal case corpora.

In [114] Hafner proposes an organization of a case law knowledge base in terms of three interacting components. First, there is a domain knowledge model that defines the basic concepts of a case law domain; these may be the actors, events and relationships from a particular domain, e.g. intellectual property or personal injury liability. Second, a legal case frame that represents the properties, roles and relationships of elements of the case and which the case instantiates. And finally, Issue/Case Discrimination Trees that represent the significance of each case relative to a model of the normative relationships of the legal domain, e.g. shephardised relationships and case-based reasoning. In current terminology, the first two would be ontologies, an ontology of the domain and of legal cases, while the last reflects the reasoning involved.

Case notes from Annotated Law Reports are used to derive the various components, among which we find:

- Legal basis: plaintiff, defendant, cause of action
- Background: description of underlying events, facts, and undisputed legal concepts.
- Issues: disputed legal questions that must each be decided.
- Procedural context: e.g. case on appeal, the lower court’s decision, and the grounds of appeal.
- Holdings: decisions on all the issues and which side prevails.

– Legal Theory: decision rules used in the case.

Cases then can be annotated with the particulars. When we search, we can search by concept, not string. For instance, we might search for cases where the defendant was a pet owner, the cause of action was personal liability injury, the background facts included injury caused by the pet, and the decision was in favour of the plaintiff.

Beyond simply annotating the particulars of the case, we must see how the particulars are tied to the decision since the cases that one wants to extract from the case base share not only the particulars, but also the reasoning from the particulars to the decision; that is, the system needs case-based reasoning. The rules that the system uses must deal with conflicting rules, exceptions, negative queries, and “customary factors that may influence an outcome” rather than necessary and sufficient factors. In [114], Issue/Case Discrimination Trees are introduced for this purpose, where there are nodes for legal issues and nodes for factors that normatively influence the decision in the case. There are several ways nodes are connected to make the system more nuanced and articulated. Finally, cases are linked to the tree, where the cases are associated with issues and factors, and the direction of the decision. The idea is that by using the Trees, one can enter the basic information about a case, then retrieve closely related and relevant cases (or vary the information to test alternative results); in other words, it uses the conceptual representation of the cases along with a system of expert legal reasoning about the cases.

[114] represents an early effort to use conceptual representations of cases together with expert case-based reasoning to support information extraction and case searching. Related work at the time and subsequently, e.g. [16, 279, 237], developed various ways of legal case-based reasoning relative to the facts and textual source. However, the conceptual representation of the domain concepts and of the case frame seem to be distinct aspects of [114] that were not taken much further in subsequent research.

One of the reasons for this, I believe, is that conceptual annotation is a knowledge and labour intensive task - the knowledge bottleneck, which is the problem of getting the textual data into a marked up form for conceptual retrieval. There is currently some progress along these lines, using well-developed, powerful text analytic tools [284]. Indeed, reviewing [114] has helped to identify some of the legal conceptual constants that can be targetted for future text analytic studies.

2.3 Trevor Bench-Capon, Gwen Robinson, Tom Routen, and Marek Sergot. Logic Programming for Large Scale Applications in Law: A Formalisation of Supplementary Benefit Legislation [33]. *Commentary by Adam Z. Wyner*

The 1980s saw the rise of one of central projects of AI and Law - the representation of law as executable logic programs. [33] describes one of the steps in the development of the project with the representation of a large portion of the United Kingdom’s legislation on *Supplementary Benefit* (SB) as an executable logical model of law in Prolog. It outlines the legislation, the task the representation supports, and the problems encountered in creating the representation.

One of the earlier, smaller stages in the project was [255], which represented the *British Nationality Act 1981* (BNA) in Prolog. More broadly, it was part of the efforts in other fields to develop expert systems and executable representations of knowledge in general and legal knowledge in particular [260, 157, 128]. The BNA study was a rather constrained, focused, feasibility study of a self-contained article of law that led to a reasonably sized application. There was no method and solutions were ad hoc.

The study with SB was intended to see what would happen with a large piece of legislation that had been revised and that interacted with other laws. It was funded by the government of the United Kingdom under a programme investigating how Artificial Intelligence software might support the application of legislation by government offices, in this case, the Department of Health and Social Services (DHSS), which was a large, complex, organisation tasked with applying social legislation. The SB itself is an extensive piece of legislation (600 pages) including definitions; aside from the legislation, there were auxiliary documents such as a guidance manual. In the end, 90% of the legislation was coded over a period of two months, though unsupervised and not evaulated.

The paper distinguishes requirements for different kinds of user of the legislation, e.g. clerks who apply the law, solicitors who advise on how the law is likely to be applied, and individuals or organisations that want to comply with the law. These classes of user do not have the same information or goals with respect to the legislation. Rather than creating different representations of the same legislation for each of these classes of user, which would lead not only to redundancy, but possible inconsistency and unclarity, it was proposed to create a logical, executable formalisation of the legislation that could be used as a core across requirements and upon which additional functionalities could be built. For example, it may be useful to provide determinations (from facts to inferences), explanations (from conclusions to their justifications), or to monitor compliance (checking that procedures are adhered to). To maintain a link between the source legislation and the code, the legislation itself is the basis of the translation to logic programming. This was in contrast to [128] that ignored the legislation itself and created an expert system, intended to mimic how a clerk might apply the legislation.

The approach was to code directly, starting top-down, based on a reading of the legislation, and without a specific methodology or discipline. High-level concepts were defined in terms of lower level concepts, and eventually grounded in facts in a database or supplied by the user. This rule-based approach contrasts with [260], which has a preliminary analysis of legislation in terms of entities and relationships. Nor was there an intermediate representation, e.g. a 'structured English' form of the legislation, which would have clarified the analysis and supported verification. Thus, complex predicates were created that could be decomposed in a variety of ways as in:

X is-not-required-to-be-available-for-employment *if*
  X is-regularly-and-substantially-engaged-in-caring-for-a-severely-disabled-person.

X is-regularly-and-substantially-engaged-in-caring-for-a-severely-disabled-person *if*
  X is-regularly-and-substantially-engaged-in-caring-for **Y**,
  Y is-severely-disabled.

There were, then, substantial issues about the formal decomposition of the predicates. The guiding principle was to decompose as little as possible.

In the course of the analysis, some general issues were identified, two of which we mention. First, a basic definition of an entitlement may have enabling provisions that specify exceptions, some of which are in statutory instruments such as regulations. Not only did such provisions have to be defined, e.g. as negations of predicates in the body of a rule, but the relation between the Act and other statutory instruments had to be coordinated, which was problematic and raised the issue of what was being represented. In practice, reference to clauses were incorporated into complex predicates, though there could be long chains of such references. An alternative, creating a database of the structure of the legislation, which would be maintained and used for reference was proposed but not tried on this project. Another general issue was representing and reasoning with temporal information, as objects and relationships change in time through the influence of events that occur in the world. For example, an individual's entitlement may change over time. In addition, there are idiosyncratic definitions in the legislation for temporal periods, e.g. what counts as a continuous temporal period may include breaks or retrospective rules, e.g. as in receiving a benefit as the result of appealing a decision. For such issues, the suggestion was to introduce some knowledge management system and the means to reason with time. Other issues were the treatment of "deeming provisions", negation, extension to case law, and open texture concepts.

This line of research, where legislation is translated into executable logic programs, has been a notable commercial success. The company Softlaw adopted key ideas and evolved to provide large scale, web-based tools to serve legislation to the public [141, 84]. In addition to executable logic, Softlaw scoped the problems, maintained links to the orginal legislative source, and added a controlled language, among other features. The company has been (after several intervening changes) acquired by Oracle, where it now provides Oracle Policy Management to governments the world over. Despite this development and success, the AI and Law research community seems not to have followed suit with similar open-source tools for research and development. In the 1990s, research interest in large scale formalisations and reuse of legal knowledge shifted to ontologies (e.g. [37], see section 7.1). For a recent legal expert system based on an ontology that can be reasoned with using description logics see [268].

2.4 Jon Bing. Designing Text Retrieval Systems for Conceptual Searching [58].
Commentary by Erich Schweighofer

Information retrieval (IR) is the challenge of using words, phrases and sentences in the best way so as to get a perfect match and retrieve only a very small number of documents from a huge text corpus. We speak of millions of documents and words using the very special legal language. The user knowledge which allows people to do this, kept and fine-tuned by professional legal researchers, is published only as a methodology [59]. For many years, IR and AI and Law researchers have tried to enhance knowledge - and computer support - for this task by working in the field called conceptual information retrieval. Some doubt that IR is hard-core AI making the life of conceptual information retrieval in AI and law not always easy. In the 1980s, rule-based approaches dominated the field in Europe (e,g, [255] and [262]). The main core of conceptual information retrieval is not matching - Boolean or non-Boolean - but analysing the use of legal language and terminology as a representation of a legal document. In the end, text, pictures, videos, concept- and logic-based representations aim for the same purpose: looking for the best and most user-friendly way of capturing the main core of a legal act, e.g. a decision of a law-maker, of a court etc. Now it is well recognised that language use remains one of the great challenges of AI and Law.

When ICAIL started in 1987, Jon Bing, Carole Hafner (see 2.2) and Joe Smith (see 7.5) represented the state of the art and the options for the future of conceptual information retrieval. Whereas Carole, as a computer scientist, choose a very formal way, Jon, as a lawyer, knew very well what it means to struggle on a daily basis with words, phrases and sentences in a very large text corpus. This may be the main difference between lawyers with practical experience (like me) and pure legal theorists and computer scientists. A lawyer has always the huge text body and his degree of mastery of a special topic in mind. For a computer scientist, a high-level formalisation with many ways of using and reformulating it is the aim. A good solution has to take account of both goals and also consider strongly both the huge size of the text corpus and the fact that it is constantly changing.

Jon's work on conceptual information retrieval in the 1980s was related to his pioneering work with Trygee Harwold on legal information retrieval, leading to the masterpiece work in 1977 [56] and the extension in 1984 [57]. At that time, the basic principles were established but the practice was still quite unsatisfactory, especially when it came to user interface, coverage of text corpora and response time.

Conceptual information retrieval has two legacies, one in the research on thesauri and the other in conceptual jurisprudence (mentioned rather less but see e.g. my habilitation thesis [252]). Jon started with thesauri as a strategy for "conceptual searching" to solve the synonym problem. In legal text corpora, an idea is represented by a large number of different words or phrases. The homonym problem was not addressed. The systems POLYTEXT[1] and LIRS (Legal Information Retrieval System) [114] developed a uniform conceptual representation for a document. The approach of Jon's department, the Norwegian Research Center for Computers and Law (NRCCL), was called "norm based thesaurus", and the area of application was the old age pension. The main innovation was the possibility of representing the structure of a document as a normalized form with arrow-diagrams (with colours), which can be seen as an early idea of the representation of hypertext structures. The representation also allows a rudimentary rule structure for the legal domain. The user can access the text retrieval system through the structure and the diagrams. Also considered was expanding search requests by additional terms (local metric relevance feedback). Terms with high co-occurrences are considered as candidates for inclusion. Despite promising results, this research, also called F*KUS concept, was never completed due to lack of resources. I pursued a similar research goal, working heavily on legal information retrieval as a way of extending the legal knowledge base (at that time mostly related to European law). I worked much more on legal language, in particular on the homonym problem [250]. Later, I used neural networks as a tool for conceptual analysis and document description [251]. In the end, I developed another model for a conceptual knowledge base [252], which like Jon's norm based thesaurus, was never implemented in practice.

Later, research on conceptual information retrieval began to merge with that on legal ontologies, as thesauri and lexical ontologies constitute a major part of this research. There is some reuse of older approaches, however, with a theoretically much stronger model focusing on legal subsumption. For some years, I have been working on a Dynamic Electronic Legal Commentary (DynELCom) [253], which provides an ontological knowledge model representing a huge text corpus and also the objects and actions in the world, offering (semi)automatic text analysis for knowledge acquisition, and various tools of legal syllogism.

Many other projects have combined conceptual information retrieval and legal ontologies. What is still missing is the scaling-up to a big application and thus a practical solution of the knowledge acquisition problem. Standard retrieval systems offer now much more support on legal language. A pop-up window with variations of the root form can be seen quite often and relevance ranking is standard now. With WIN, a very fine but not explicit representation of legal language exists. What is missing, however, is a Google-like efficient ranking

[1] Joint research project of the Stanford Research Institute, California and the KVAL Institute for Information Science, Stockholm.

of documents and some harmony between linguistic support, Boolean search and ranking. So far, expert users seem to be slightly uncomfortable with intelligent solutions [170].

It may take some years before we get ontological (and thus conceptual) representations of legal text corpora. The decisive question is whether legal writers are really considering moving to a new form of representation and using also (semi)automatic text analysis. Then, such a representation could greatly improve retrieval of relevant cases and ranking of results. Eventually, legal research may be supported in a similar way to that Google provides for some other areas today.

3 Vancouver 1989

The Second ICAIL was held June 13th to 16th 1989 in Vancouver, on the beautiful campus of the University of British Columbia, with Joe Smith and Robert Franson as Conference Co-Chairs. Edwina Rissland was Programme Chair and the Programme Committee had expanded to eight, now including three Europeans (two UK and one Norwegian) and five US-based members. Roger Schank was an invited speaker, and the salmon barbecue held outside amongst the totem poles of the Museum of Anthropology at UBC was a memorable highlight. The panel this time concerned research funding, especially its opportunities and pitfalls. Several papers describing a range of expert systems were presented, and the four papers selected in this section reflect different approaches to modelling legal knowledge: the pure case-based approach of HYPO (3.2), the production rule approach of EPS (3.3) and the hybrid approach of CABARET (3.4). The logic approach, including the necessity and desirability of deontic logics and deep models was discussed in [27] (3.1).

3.1 Trevor Bench-Capon. Deep Models, Normative Reasoning and Legal Expert Systems [27]. *Commentary by L. Thorne McCarty*

This contribution by Trevor Bench-Capon [27] was primarily a position paper, but it is significant because it introduced into the AI and Law literature a concept that subsequently played a prominent role: *isomorphism*.

The position taken by the paper was unexceptional: at the time, there was much discussion over the relative merits of "shallow" rule-based expert systems, e.g., MYCIN [256] and PROSPECTOR [85], as opposed to systems that were based on "deep conceptual models" of the relevant domain [176]. In the legal realm, these deep models would surely include a representation of the deontic modalities [175], but other modalities (temporal, causal, epistemic, etc.) would also be important. Trevor's main point was simply that the resolution of this question depended on the level of adjudication at which the expert system was being applied. For "a higher level adjudicator confronted with a case in which the law is not clear [p. 43]," deep models and normative reasoning would probably be necessary. But there are many low-level adjudicators: "they abound in government departments administering welfare benefits, tax and other laws [p. 44]," including "such quasi-legislation as the internal regulations of companies [p. 44]." For these routine legal decisions, Trevor asserts, deep models and normative reasoning are "neither necessary nor appropriate [p. 44]."

Put this way, who could disagree? The novelty of the paper, however, lies in its characterization of legal expert systems in a way that distinguishes them sharply from "shallow" expert systems in fields such as medicine and geology. In the law, Trevor writes, an expert system can be based on a formalisation of the legislation itself, and a faithful representation of the legal rules can provide many of the benefits claimed elsewhere for deep models. Here is where the concept of *isomorphism* was first introduced, albeit obliquely. The term "isomorphism" appears only four times in the paper, and three of these occurrences are in negative contexts, but the one positive reference is the following: "Isomorphism remains, however, the aspiration, and the more faithful the formalisation, the greater the advantages of the approach [p. 39]." Once this concept has been articulated, of course, the term develops a life of its own. A more definitive paper on "Isomorphism and Legal Knowledge Based Systems," written by Trevor with Frans Coenen, and published in 1992 in the very first issue of *Artificial Intelligence and Law* [35], cites the ICAIL 1989 paper as one of the two earliest references on the subject,[2] but there were several more.

Is *isomorphism* still a relevant concept? Twenty years after the ICAIL 1989 paper, Trevor and Tom Gordon published a paper in ICAIL 2009 entitled "Isomorphism and Argumentation" [43], which brings the analysis up to date. In the intervening years, argumentation theory has been applied extensively to the "rule-plus-exception" structure that is typical of most legislative drafting, and a theory of isomorphism today must take these developments into account. As a case study, the 2009 paper analyzes a fragment of German Family Law, and presents four possible representations, three of which are isomorphic! The lesson seems to be that isomorphism is still an important touchstone, but other considerations (such as the allocation of the burden of proof, and the efficient operation of an administrative system) may cut in different directions. Finally, there is a curious section at the end of this paper on the "Role of Ontologies." Here, an alternative representation of the Family Law fragment is discussed, based on a description logic that is coupled with the qualifiers: *allowed* and *disallowed*. Although Bench-Capon and Gordon are quite critical of this alternative representation, its very existence suggests a subversive question: An ontology is, first and foremost, a "deep conceptual model" of a particular domain, and the qualifiers "allowed" and "disallowed" are simple deontic modalities. Is it possible that a thorough treatment of isomorphism, today, would lead us back to both deep models and normative reasoning, even for low-level adjudication?

3.2 Kevin D. Ashley. Toward a Computational Theory of Arguing with Precedents [15]. *Commentary by Henry Prakken*

In May 1988 I started a four-year PhD project at the Computer Law Institute of the Free University Amsterdam. I had studied law and philosophy in Groningen and I had just completed my MPhil thesis about legal applications of deontic logic. My PhD project was titled *Logical aspects of legal expert systems* and the project description left me with almost complete freedom what to do. Naturally I thought about continuing my MPhil research as my PhD research.

However, in my first month as a Phd student our institute was visited by Edwina Rissland. As preparation I was given to read all the papers then available about HYPO. At some point I was given the opportunity to talk to her. She asked me what I wanted to do and I could think of no more than summarising my MPhil thesis and uttering some vague ideas about continuing to work on it. I saw her thinking "this is leading nowhere". Maybe she still thinks that my research has led nowhere but to this day I am grateful to her for saving me from writing my thesis about deontic logic. I found the papers on HYPO fascinating and

[2] The other early reference is to [148], which does not seem to be available online.

they fully convinced me that the central theme in modelling legal reasoning is argument. However, logic-minded as I was, I was puzzled by the relation of HYPO to logic. Fortunately, one month later I attended a tutorial by John-Jules Meyer, which introduced me to the field of nonmonotonic logic. This planted the idea in my head that legal argument could be formalised as a kind of nonmonotonic logic.

My Phd thesis, which I defended in 1993, presents my first attempt to formalise this idea (not good enough to be remembered), but I did not yet address HYPO-style legal reasoning. This came only after I had published my first decent work on logics of argumentation, with Giovanni Sartor in [209,210]. After finishing this work, we turned to the formalisation of HYPO-style legal argument with our newly developed logical tools for argumentation, resulting in [214,212]. One of our key ideas was to represent a precedent as two conflicting rules *pro-plaintiff factors* $\Rightarrow p$ and *pro-defendant factors* $\Rightarrow \neg p$ (where p represents the decision that plaintiff wins) plus a priority: if plaintiff won then the first rule has priority over the second, if defendant won than the second rule has priority over the first. (This way of logically representing cases was later adopted by e.g. [41,78,135]). We then modelled various case-based argument moves as moves in a formal dialogue game that presupposes our [210] argumentation logic.

An initial problem for us was that the first publications about HYPO, although exciting and fascinating, were too informal for logic-minded researchers like us. This is where Kevin Ashley's ICAIL 1989 paper turned out to be a great help, since it is the first paper that gives clear formal definitions of the kind of reasoning implemented in HYPO. It uses some elementary logic and set theory to define the notions of relevant similarities and differences between cases and analogous precedents, and then to define various roles of precedents in legal arguments, such as a cited case, a distinguished case, a (trumping or non-trumping) counter example and a target case for hypothetical modification.

To be honest, Ashley 1989 was not our only source of inspiration, since there had been other papers of logic-minded AI and Law researchers in the same spirit, notably [116,166, 167]. However, for me personally it was above all Ashley's paper that made made me believe that case-based legal reasoning can be logically formalised.[3] More generally, the ideas expressed in this paper have, I believe, played an important role in bridging two research strands in AI and Law. At the time of publication of Ashley's paper, there was still a strong debate between proponents of "case-based" and "rule-based" models of legal reasoning, where some at the case-based side seemed to believe that the use of logical methods inevitably commits to a rule-based model. However, as papers like [41,240,78,135,285] show, it is nowadays widely accepted in our community that logical models of legal reasoning with cases are both possible and worthwhile.

3.3 D. A. Schlobohm and L. Thorne McCarty. EPS II: Estate Planning with Prototypes [249]. *Commentary by Kevin Ashley*

This paper appeared in the heyday of legal expert systems development. At the time, as the authors point out, AI and Law researchers focused primarily on developing systems for legal analysis using heuristic production rules and, to a lesser extent, logical models of statutes and case-based techniques. In this paper, however, Schlobohm and McCarty focused

[3] To my embarrassment I must say that until now I never cited Ashley's 1989 paper, since I always cite his book on HYPO [16]. Most other authors do the same, which explains why his 1989 paper is not much cited. However, I have always used not the book but my hard copy of the ICAIL 1989 paper to check Ashley's definitions, witness the many handwritten notes it contains.

squarely on the important task of legal planning: “the principal activity of business and tax attorneys is not litigation, but rather assisting clients in planning their affairs so as to avoid litigation in the future.” The relative dearth of AI and Law work on legal planning, of course, has continued to the present day.

Dean Schlobohm had constructed a heuristic rule-based expert system (EPS) to analyze clients’ needs and recommend a testamentary estate plan. (Schlobohm still practices estate planning in San Francisco.) The aim of this paper, however, was to incorporate into a proposed EPS II, a “deep model,” in Thorne McCarty’s sense of the term, of the legal estate planning domain, including the relevant tax codes, and use it to construct a set of estate plans that satisfy a client’s goal.

The work was not just focused on legal argument or statutory inference but on a creative planning task for which arguments and statutory inferences are relevant but not ends in themselves. The planning task connected to a “tradition” of classic AI planning, but a key insight of the paper was to employ *prototypes*, prototypical estate plans used as exemplars that could be matched to a client’s situation and *deformations*, transformations of various modifications in the estate plans that preserve or defeat the client’s goals. Thus, to make the planning task tractable, the authors proposed to use a kind of case-based AI approach, complete with goal-based adaptation. Given the overall estate planning goals of providing for beneficiaries, while retaining client control over assets and avoiding taxes as much as possible, the authors identified a number of prototypical estate plans and the transformations that could render them more effective in achieving the right goal tradeoffs. For example, the Clifford Trust prototype created a ten-year-plus trust for a child’s educational benefit without a permanent transfer of assets but a subsequent tax reform rendered it much less effective. The “pipe dream” prototype maintained the client’s control of his assets but at a high cost in taxes in light of certain tax code provisions. Given a client’s particular circumstances, however, various transformations of the “pipe dream” prototype could accommodate the tax provisions and the client’s objectives.

The intriguing challenge was how to get a computer program to retrieve a prototype plan as well suited as possible to a client’s needs and adapt it with reference to the tax code constraints in order to maximize the client’s goals. The deontic machinery of McCarty’s Language for Legal Discourse (LLD) [177] was invoked to represent the permissions, obligations, and things “not forbidden” under the tax codes and his theory of prototypes and deformations in TAXMAN II [174] to model the transformations. While in TAXMAN II, legal concepts were represented as a set of (sometimes prototypical) exemplars, here, particular estate plans became the exemplars and the transformations were various modifications in the estate plans that either preserved or defeated the three goals as they pertain to a particular client. Importantly, presumably through the mechanics of LLD, the allowable transformations would be only those that preserved the conceptual coherence of the tax code’s estate planning regulations.

While I did not then, and do not now, fully understand the deontic mechanisms of LLD that would preserve conceptual coherence in the transformations, I remember being fascinated by the analogy (at the time I might have called it a “sleight of hand”) that linked legal concepts in Taxman II with legal plans in EPS II, but that is a connection that has endured. Legal plans, like legal concepts, are artifices designed to achieve or accommodate certain legal and practical goals. The legal concepts are attempts to achieve certain legislative purposes underlying the regulatory texts in which the legal concepts are embedded. The legal plans attempt to accommodate the meanings and purposes of those legal concepts in the context of the regulatory texts. Legal plans and concepts both connect to and interact with the texts and the logical rules embodied therein, but they also share the remarkable property of plasticity: they can be transformed in ways that interact with the texts and purposes, trading off some goals over others, with results that can be more or less effective, coherent, reasonable, and defensible.

This is an important lesson every law student needs to learn, and therein lies another reason why I regard this paper as remarkable. In a lifetime of law teaching, one encounters relatively few pedagogically useful extended examples of creative lawyering, and this paper provides several. Constructing a legal plan and adapting it to a client's constraints is a key component of the practice of law and yet another skill that law schools do not teach very well. Can one build a program that lets students play with the construction of legal plans, modifying them to achieve certain practical goals, and experiencing the consequences given a regulatory scheme? Or modifying the regulatory scheme in various ways to experience the consequences for accepted legal plans? By "experiencing the consequences" I mean in terms of modeling the effects on the explanations and justifications of a plan. The program that powers such a tutoring environment need only work for a half a dozen plans and a small set of reasonably complex regulatory provisions. If it could draw on a deep model of the regulatory domain, thus avoiding "some of the brittleness which traditional expert systems experience when fact situations approach the limits of the system's knowledge," it could even deal gracefully with students' impulses to drive the plans off the nearest cliff. An EPS II that did little more than this would accomplish a great deal. In a realistic tutoring context, it would enable intriguing empirical studies of how to teach legal planning as a case-based design task. It would also enable some AI and Law researcher to pick up the lost trail of computationally modeling legal planning that Schlobohm and McCarty blazed in this paper.

3.4 Edwina L. Rissland and David B. Skalak. Interpreting Statutory Predicates [228].
Commentary by Ronald P. Loui

David Skalak and Edwina Rissland, in their modest CABARET papers at ICAIL [228] and IJCAI [229], provided a lesson in semantics for me that Wittgenstein and Waismann were unable to provide. Sometimes it takes a single clear example, with a memorable diagram, to turn the light on. For me, that light was Rissland's diagram showing the tree of conditions for arguing the tax-deductibility of a home office.

At the top of her diagram were the few necessary conditions:

1. regular use;
2. exclusive use;
3. principal place of business;
4. if an employee, then some other conditions.

As her tree descended, she explained that some nodes could have other expansions, such as rules for daycare providers or for storing inventory, and that at the leaves, case-based argument would be required to pin down the open-textured terms. What makes something the principal place? What counts as business activity?

It was a direct depiction of H.L.A. Hart's view [129], which I had not heard of in 1989 despite a broad exposure to philosophical logic, and in particular, non-demonstrative reasoning. The "UMASS" model joined the case-based mechanics from Rissland and Kevin Ashley's HYPO [227] and the search-tree diagram from automatic theorem-proving (as one might find it in a textbook on AI, such as Nilsson [186]). We had been drawing similar diagrams since about 1985 for knowledge representation. Our diagrams were supposed to be simplified ways of reasoning epistemologically, or ways of jumping to conclusions about knowledge representation shorthands. They were not supposed to be a theory of meaning for constructible properties.

These days, whenever I consider of the meaning of non-scientific language, I immediately conceive in my mind the picture that Rissland projected during her talk. Often, I am frustrated that other people have never seen such a picture. Very well trained philosophers of language seem to have no problem adopting this picture, even though I am sure they have never seen Skalak-Rissland's papers, or even Hart's motivating work. But they usually must be very mature in order to have this view. While the picture fits directly within the Wittgenstein semantic traditions (Hart being enamored of Wittgenstein), it is a far leap from the *Zettel* of the great philosopher to the practical clarity of the Rissland-Skalak tree.

Rissland-Skalak is only a small departure from HYPO, emphasizing the rule-based structure on top of the case-based reasoning. But the superstructure is crucial. In their work, it is clearly an analysis of meaning. In HYPO, the domain-specific details obscure the greater lesson. Also, HYPO is presented mainly as a normative argument-system for adversaries, or a simplified descriptive model. One might explain that the procedural nature of fixing the meaning of terms, such as legal terms, is tied to this adversarial argument. But if one has to explain, then one probably has not found the clearest example.

Rissland-Skalak also motivated my work with Jeff Norman ([166] and see 5.5) on the argumentative structure of cases. Rather than declare a level at which case-based reasoning becomes an accounting of pro and con factors, we grow the argumentation tree further, asking the case-based reasoning to consider the prior arguments used in the prior case. Giovanni Sartor and Henry Prakken gave a paper along similar lines at ICAIL in Melbourne [211].

When I attended a University of Chicago law school symposium on AI and Law, Cass Sunstein complained that HYPO seemed too small a piece of the puzzle to be the full model of the analogical reasoning required for open-textured terms. I felt at the time, and still feel, that if he had seen the later developments of the UMASS legal CBR, Skalak, Loui-Norman, Sartor-Prakken, and others, then he would have left the symposium enlightened. Incidentally, that was the symposium where students of law lecturer Barack Obama suggested that their teacher would make a great national political candidate, which set in motion the future U.S. President's invitation to speak at the 2001 ICAIL St Louis (state senator Obama later withdrew as speaker because of the legislative schedule).

A model of legal meaning is important because most of the language one encounters in the world, and most of the practical uses of language used in databases and knowledge bases, is constructible and open-textured. That means that the extension of predicates is subject to "lazy learning," not determined in advance by a set of rules, but determined as cases arise, and as those cases are decided according to meta-rules. It is the kind of meaning that one encounters in human-human dialogue, human-machine dialogue, in most areas of human experience.

The other successful examples of modeling meaning are found in mathematics, science, and engineering. Although these human modes of inquiry (and control) are the inspiration of well-mathematized models, well-promulgated and well-understood, they are not the models that are applicable most of the time. Most of the time, it is a process-based model of constructible meaning that is applicable, and the paradigmatic example is legal language. So we have analytic meaning, in which necessary and sufficient conditions are in play, and things like fuzzy and distributed or activation models of meaning. Familiar metrical mathematics appears in those models, and they are seductive to anyone who likes to play with those metrics. But the discrete-mathematical models of rule-based defeasible argument and case-based analogical reasoning are also precise and support a kind of intellectual play. Ironically, those who developed them have possessed considerable mathematical backgrounds.

(Ironically, Edwina Rissland once accused me of liking an approach just because it had more mathematics in it. My retort was that she herself had been trained in mathematics. That was one year before my Rissland-Skalak epiphany.)

Consulting at Cycorp, I have recently found a delightful collection of well-trained (most have Ph.D.'s) philosophers of language and philosophers of logic who have been tasked with the creation of a rule base and knowledge base that actually does things. There are numerous classical issues that arise, most of which have been argued and turned into best practices. Unlike an academic community, the advocates on each side of an issue do not continue to publish on the topic for years. Sometimes there is no clearly superior point of view, and pragmatics simply requires that an arbitrary choice be made. There is often not much time to dwell on issues once the ramifications have been determined. So one finds a Davidsonian approach to events, a TMS method of belief revision, and various kinds of default, defeasible, and meta-reasoning. People don't tend to make a big deal about the schools of thought that Cyc represents in its architectural choices.

There is not yet a leaf-layer of case-based reasoning by analogy that grounds open-textured terms in Cyc. But the constructive meaning by continuous development of predicates is an unavoidable feature of KB engineering. As Cyc is an on-line reasoning artifact, changing day-by-day as multiple users add and modify rules, people are forced to view meaning as a dynamic phenomenon. A predicate's meaning is necessarily imperfect (corrigible/revisable), and necessarily usable in its imperfect states (anytime/online). This is not quite the Skalak-Rissland view of Hart, but it feels familiar.

I wasn't planning to attend the Skalak-Rissland talk at IJCAI. What I knew about AI and Law, from Thorne McCarty, was that there was such a thing as non-probabilistic non-demonstrative reasoning. But I wandered into the talk and sat down, and have been an advocate of AI and law, legal language as a central model of language, and process-oriented open-textured meaning ever since.

4 Oxford 1991

The Third ICAIL moved to Europe for the first time: it was held June 25th to 28th, 1991 in St Catherine's College, University of Oxford. Richard Susskind was Conference Chair and Marek Sergot was Programme Chair. The Programme Committee was now up to ten, divided between five Europeans, all from different countries, and five North Americans including one Canadian. The Conference was opened by the Lord Chancellor, the Right Honourable Lord Mackay of Clashfern, and Neil MacCormick, a distinguished Scottish jurisprude was an invited speaker. There was a panel on rules and cases, chaired by Don Berman. Several papers concerned various software engineering aspects of legal expert systems, two of them noting the rapidity with which the law changed, necessitating maintainable systems. Also interesting was [141]: after the conference the authors set up SoftLaw, a company which was to become the most successful commercial spin-off applying legal expert system techniques, and which is now owned by Oracle. Three papers have been chosen from this conference. Two take up themes from the paper discussed in 3.1: whether we need to model substantial knowledge of the world (4.1) and whether we need to model norms and use deontic logic (4.2). The paper discussed in 4.3 introduces themes relating to argumentation that are still discussed: what moves are available and can we describe strategies for deploying them?

4.1 Joost Breuker and Nienke den Haan. Separating World and Regulation Knowledge: Where is the Logic? [70]. *Commentary by Trevor Bench-Capon*

It has taken me a long time to understand this paper, and it may well be that I do not fully understand it even now. At the time I thought it was about a kind of legal expert system: an interesting application. As time has passed, however, I have come to see that it addressed some rather more profound questions, questions which are central to AI and Law, and to argumentation, and which are relevant today as they were in 1991.

Viewed as an expert system the paper describes a system, TRACS (Traffic Regulation Animation and Comparison System) which when presented with a traffic situation determines whether the norms codified in the regulations are violated. But what is key here is the knowledge representation, in particular the separation of the *world knowledge base* from the legal knowledge contained in the *regulation knowledge base*, and the knowledge of how to resolve conflicts in the legal knowledge held in a third knowledge base. What is at stake here is how we can start from a set of facts and end up with a legal conclusion. This the old question posed by David Hume [136]: *how can we derive ought from is?*. That this is a problem is clear from a classic legal knowledge based system based on a formalisation of legislation, such as the famous British Nationality Act system [255]. Although in that program, which is intended to contain only a formalisation of regulations - legal knowledge - it appears we derive that Peter *is* a British citizen from the facts that the father of Peter *is* John and John *is* a British citizen, in fact our conclusion is normative and our facts are also legal facts: biological parenthood is no determinant of the truth of fatherhood for the purposes of this system.

The issue is clearly stated in a series of papers by Lars Lindahl and Jan Odelstad, of which [160] is representative. They focus on the notion of *intermediate* predicates which bridge from facts to normative consequences:

1. $Facts \rightarrow Intermediates$
2. $Intermediates \rightarrow LegalConsequences$

The problem is that while the reasoning associated with step 2 can be relatively clear, the reasoning associated with step 1 is far more mysterious. Sometimes step 1 uses explicitly stated rules, but even here the rules are defeasible and subject to interpretation and exceptions. Worse, as discussed in [20], sometimes there are no rules, but only a set of features of the situation to be considered, none necessary and none sufficient. This relates strongly to the notion of *factors* as found in CATO [5] and IBP [18], which can themselves be seen as intermediate predicates. The problem is that while AI and Law has been quite successful at modelling the reasoning involved in step 2, using a variety of techniques, some involving case-based reasoning, but also using rule-based systems and description logics [268], it has made far less progress in the more difficult task of modelling step 1. Once we reach step 2 logic can take over, and computational support becomes very possible, but step 1 seems to defy such automatic processing. But if we cannot replicate step 1 we cannot capture the whole process: we cannot move from a set of facts to a decision.

Part of the problem is that world knowledge is so extensive. Breuker and Den Haan quote Thorne McCarty as saying that it would need to encompass "space, time, mass, action, permission, obligation, causation, intention, knowledge and belief and so on", and point out that these categories would require several specialised ontologies. The ontological requirements to support such reasoning were more recently discussed in [17] (see 13.1). But these requirements are enormous, and very often specific to particular cases. As Thorne argues in 13.1, it is highly unlikely that this degree of modelling can be done on a scale which will enable practical support. So perhaps we should keep world and regulation knowledge very firmly separate, and focus our systems on performing the relatively straightforward task of reasoning with regulation knowledge. Capturing in a computational system the world knowledge required for analysis of cases into intermediate predicates may forever remain beyond us. We can still build useful, supportive systems, but representation as intermediate predicates will have to be done, as now, by skilled human individuals.

4.2 Henning Herrestad. Norms and Formalization [132]. *Commentary by Guido Governatori*

The starting point of the paper is the question whether deontic logic is needed for the representation of legal knowledge in Artificial Intelligence. At the time the paper was written a prominent view, supported by some members of the logic programming group at Imperial College was that legal knowledge is mostly definitional in nature and thus it can be represented by sets of logic programming clauses without the need of deontic operators (e.g. [27] and see 3.1). Herrestad argued against this position.

The analysis is based on a classical problem in deontic logic and normative reasoning: the formalisation of the Chisholm paradox [77]. Shortly the paradox is about a set of sentences in natural language where, intuitively, the set is consistent and the sentences are (logically) independent from each other. The crucial aspect of the Chisholm set is that it includes factual statements and conditional normative statements, in particular an obligation in force as a response to a violation whose obligation is expressed by one of the other sentences. These constructions are known as Contrary-to-Duty obligations or imperatives.

The paper discusses the formalisation of the paradox proposed by Sergot [254], and argues that while it is consistent it fails to satisfy the independence criterion. This can be seen from two perspectives: the first is that the formal representation is not a faithful translation of the sentence, which somehow can be seen as a matter of interpretation of legal texts. The second is about the logical consequences of this, namely, when a knowledge base is queried about a counter-factual, the knowledge base will return an unlimited number of wrong answers.

Solutions to the Chisholm paradoxes have been proposed based on the idea that the obligations the sentences refer to are relative to instants and periods of time. Herrestad agrees that time is essential for a proper representation of norms, but it is not a key factor for the solution of Contrary-to-Duty paradoxes (and atemporal versions of the paradoxes have been provided).

The final part of the paper is dedicated to presenting the Deontic Logic of Ideality and Sub-ideality of Jones and Pörn [142, 143] and how this logic addresses the issues of Contrary-to-Duty obligations in a simple and elegant way.

The major contribution of the paper is that is shows a non deontic logic approach to the formal representation of norms is not appropriate, in particular in the presence of norms providing compensations for violations (this is also true for approaches based on deontic logic based on standard normal modal logics, e.g., modal logic KD corresponding to the so called von Wright's Standard Deontic Logic). Herrestad concluded his paper expressing the belief that there are many tasks and areas of law requiring genuine deontic reasoning. In [107] I provided support to this belief by pointing out that contracts contain clauses describing the obligations and other normative positions for the parties involved in a contract as well as compensatory clauses (i.e., contrary-to-duties obligations). Thus contracts are a natural arena for practical deontic logic. In the recent years we have experienced a renewed interested in the formalisation of norms (e.g., applications for service level agreements, electronic contracts and regulatory compliance). Several approaches advance potentially efficient computational methods based on either non deontic approaches or basic deontic logic frameworks, thus ignoring the lesson taught by Herrestad.

The paper left an open question of whether it is possible to give automated reasoning techniques for the logic of Ideality and Sub-ideality of Jones and Pörn. I took up the challenge and provided a positive answer by developing a labelled tableaux for the logic [106].

4.3 David B. Skalak and Edwina L. Rissland. Argument Moves in a Rule-Guided Domain [257]. *Commentary by Katie Atkinson*

Arguments play a central role in legal reasoning. Their study in relation to AI and Law covers many different aspects that include, amongst other things, explanation of legal decisions, the structure of arguments in legal cases, and the role that arguments can play in case-based reasoning. Skalak and Rissland's focus in their 1991 ICAIL paper [257] was concerned with the important issue of *strategies* for creating arguments; selection of a good strategy is needed to optimise the likelihood of being persuasive, and of the agent putting forward the argument being successful. The backdrop of this work on argument strategies is complex domains where a rule-based representation is employed and the aim is to use previous cases to formulate arguments that convince of a particular interpretation of a rule in a new fact situation. The problem arises from a host of issues with the representation of rules, including the use of terms that are not well-defined, exceptions that are not made explicit in the formulation of the rule, and enthymematic issues. Legal statues can be represented in the form of such rules and are thus used in [257] as a domain in which the development of such argument strategies can be considered.

The strategies that are proposed for arguing about rules take into account the point of view of the arguer by characterising whether she is taking a *pro* or *con* stance on the issue, i.e. her goal, and her interpretation of the rule being considered. The strategies are intended for use in scenarios where an initial determination has already been made as to whether a rule's conditions have been met and a respondent is arguing about whether this determination should be applied. This characterisation yields four strategies: *broadening*, where the stance is pro and the conditions are not met and it is proposed that the scope of the rule's application should be widened; *confirming the hit*, where the stance of the arguer is pro and the conditions are met and it is proposed that the rule should apply in the new situation under consideration; *confirming a miss*, where the stance is con and the conditions are not met and it is proposed that the rule should not apply to the new situation; *discrediting*, where the stance is con and the conditions are, strictly speaking, met but it is proposed that the rule should not apply to the new situation under consideration.

The above strategies are effected through *argument moves*, whereby for each strategy there are four possible moves. These moves are determined by considering, firstly, the disposition of a precedent – whether or not it held for the desired result, and, secondly, the status of the rule with respect to the precedent – whether or not the rule's conditions were met by the precedent's facts. The argument moves themselves involve the general argumentative tasks of *analogising* and *distinguishing*, which the authors characterise as *primitives*.

Given the above characterisation of the argument strategies, moves and primitives, arguments themselves can then be formed. To do this, a decision tree can be constructed that maps out all the elements of representation, as described above, enabling the tree to be traversed in a top-down manner to construct an appropriate argument for a given situation.

The final contribution of the paper is to take the argument strategies, moves and primitives and implement them in a computational system. For this, the authors use their CABARET system [230], which is an architecture that makes use of rules and cases to solve case-based reasoning problems. The performance of systems that make use of production rules is very much dependent upon the control structure that is used to determine the order in which rules fire. As such, the theory of argument strategies and moves set out by the authors is used as the heuristic in the control structure of the CABARET system to generate the variety of arguments that the underlying representation yields.

Although the importance of representing and reasoning with arguments had previously been well recognised within AI and Law, the Skalak and Rissland paper was the first to explicitly tackle the issue of argument strategies for use with reasoning about the application of statue rules, and specify this in a way to enable the representation to be realised in a computational system.

Subsequent to the ICAIL 1991 paper, and the journal articles on CABARET [230] and [258], Skalak and Rissland went on to develop their work in the BankXX system; (see [234] and 5.3). In this system, specifically instantiated for use in the area of personal bankruptcy, argument generation is considered from the bottom-up perspective, as opposed to the top-down approach employed in the CABARET system. The bottom-up perspective enables a structured collection of information to be explored for material that can provide the foundations for an argument, which is an approach that the authors have found to occur frequently in practice since a clean top-down specification is not always at hand. The BankXX system went beyond the work on CABARET to bring together aspects of case-based reasoning, heuristic search, information retrieval and legal argument into a unified system. An extensive series of experiments was conducted to evaluate the program [235] and these yielded positive results, cementing this thread of work on strategies for legal argument within the notable contributions made to AI and Law research.

Outside of Rissland and Skalak's own line of research, the work has been built on by other AI and Law researchers. Amongst other contributions, Bench-Capon and Sartor's work on theory construction in case-based reasoning presented in [41] has been influenced by the types of argument move described above, as has Prakken and Sartor's formal dialogue game for legal reasoning with precedents [211]. The enduring importance of the work can be seen further in more recent publications such as [285] wherein pattens of arguments that capture the kind of reasoning in Skalak and Rissland's argument moves and strategies are characterised as argumentation schemes for case-based reasoning. As the first paper to popularise the importance of characterising legal argument in terms of moves in a dialogue, and the process of constructing arguments as the strategic deployment of these moves, ideas which since have become part of the accepted fabric of AI and Law, the ICAIL 1991 paper represents an important milestone in the development of the field.

5 Amsterdam 1993

The Fourth ICAIL was held from 15th to 18th June 1993 in Amsterdam at the Free Univerity (Vrije Universiteit) of Amsterdam. Anja Oskamp was Conference Chair and Kevin Ashley was Programme Chair. The programme Committee was now eleven: six Europeans from five countries, four North Americans, including one Canadian, and, for the first time, an Australian. Among the invited speakers, Lothar Philipps gave an interesting address (see [195]) in which he presciently introduced the AI and law community to game theory, now so popular in multi-agent systems. Drinks were held on a boat cruising the canals of Amsterdam taking us to the destination for the Conference Dinner. No fewer than seven papers have been chosen from this conference, reflecting the number of new ideas, relating to dialogues (5.4), argumentation (5.3, 5.4 and 5.7), and the rationales of decisions (generally in 5.2 and 5.5 and specifically in terms of purposes in 5.1). There were also a number of different techniques on display in the papers presented at this conference, including heuristic search (5.3) and neural networks (5.6).

5.1 Donald H. Berman and Carole D. Hafner. Representing Teleological Structure in Case-based Legal Reasoning: The Missing Link [46]. *Commentary by Trevor Bench-Capon*

This paper forms part of a trilogy of papers by Don Berman and Carole Hafner looking at various aspects of case-based reasoning. In 1991 they had examined the influence of procedural context and in 1995 they would look at signs that a settled doctrine was about to change, but here they examined the influence of underlying social purposes on decisions in hard cases. After Don's untimely death, Carole consolidated these papers into a journal article [115]. Don presented this paper in Amsterdam in his usual highly entertaining style: the cases were amusing and Don played up these aspects. But there was a highly serious point: what do you do when the cases run out: in other words, how do you decide between two sets of precedents when they favour opposing parties, and both can be distinguished from the current case? Their answer was to consider the *purpose* the decision will serve: so in deciding the well known property case of *Pierson v Post*[4], the court had to choose between maintaining the clarity of the law and encouraging a socially useful activity, and chose the former value over the latter.

The influence of the paper was not immediate. In the years that followed there were attempts to represent case-based reasoning using rules and defeasible logic (e.g. [212]), but how to go beyond conclusions which could be drawn *a fortiori* on the basis of precedents (as in [120] (see 9.3)) was not yet well understood. By 1999 I had become interested in Perelman's notion of an audience [194], and was developing a characterisation of audiences in terms of an ordering on social values (most fully described in [31]). Recalling Don's presentation and attracted by the fact that the three cases represented the leisure pastimes of an English gentleman (huntin', shootin' and fishin'), I reread [46] and found it gave me the answer I was looking for. We can extend the application of precedents by identifying the *value preferences* they express and then applying these value preferences to different sets of factors representing the same value preference. An early version of this idea was published in [30], together with responses by Giovanni Sartor [244] and Henry Prakken [204]. The ideas here were further developed and led to [41]. The notion of purpose, and with it the notion of audience, is a powerful way of *explaining*, rather than simply noting, differences of opinion, differences between jurisdictions and differences from one time to another in terms of the surrounding social context. Thus as social values change, so can legal decisions[5]. Moreover, it gives a basis for arguing why one factor *should* be preferred over another, and so justifying choices.

Following this resurgence of interest in [46], the three "wild animals" cases, sometimes augmented by additional cases, became a very popular example in AI and Law. Recently the case of the fugitive baseball in *Popov and Hayashi* has also been included in this line of

[4] Post was chasing a fox with horse and hounds. As he closed in for the kill, Pierson killed the fox with a fence pole and bore it off. Pierson won since Post did not have clear bodily possession of the fox, even though this might discourage fox hunting.

[5] *Stare decisis* must bow to changing values, as Justice Marshall put it in *Furman v Georgia*.

research. The notion of purposes or values had a profound influence on my research: I went on to pursue value based argumentation generally with Katie Atkinson (née Greenwood) starting with [110] (see also 10.2), and empirically investigate the ideas of [41] with Alison Chorley (e.g. [78]).

More widely the notions of purpose, teleology and value are of considerable current interest: see for example [246] and [109]. How values should be deployed in legal argument is still unclear: [32] recently posed the following questions:

1. Is promotion and demotion of values to be regarded as boolean (either promoted or not), ordinal (instances of promotion can be ordered), qualitative (instances of promotion can be placed into bands), or quantitative (instances of promotion can be assigned specific numeric values)?
2. Should we see promotion and demotion as relative to thresholds, or to be considered as trading-off against one another?
3. Should values be considered separately, pairwise, or collected together into complete sets?

Thus [46] is the source of a powerful and important strand of modelling legal argument, the use of purpose and value in justifying legal decisions. It can be seen the fountainhead of work in AI and Law which invigorated the case base approach to legal reasoning, and helped to bring case-based and rule-based approaches into a common framework centred on theory construction. Finally the issues raised in [46] remain an important source of current research questions.

5.2 L. Karl Branting. A Reduction-Graph Model of Ratio Decidendi [67]. *Commentary by L. Thorne McCarty*

There are two main ideas in this paper [67] by Karl Branting, and both are mentioned in the title: “reduction-graph model” is a concept from the field of computer science, and *ratio decidendi* is a concept from the field of jurisprudence. The paper is significant because it combines these two ideas to support a claim about the adequacy of a pure exemplar-based theory of legal reasoning.

Computer scientists, especially those familiar with logic programming, will recognize Karl’s reduction-graph model as a variant of an *SLD-derivation*. (For example, compare Figure 2 in [67] with Figure 1 in [161]) What Karl refers to as a *legal warrant*, i.e., “a proposition expressing the conditions under which a legal predicate is satisfied [p. 42],” is actually a *definite clause* in Horn clause logic. What Karl refers to as *exemplars*, i.e., “collections of facts, expressed in a concrete case-description language [p. 43],” are actually sets of *unit clauses* in Horn clause logic. Finally, a *reduction operator* [p. 43] corresponds to a single step in a resolution proof, in which an *atom* in the *body* of one clause is resolved with the *head* of another clause. String a sequence of these reduction operators together, and you have an SLD-derivation, or an *SLD-tree*, and Karl argues that this tree should be viewed as the “justification” of the legal decision. “A justification for the conclusion that a predicate applies to a case therefore consists of a warrant for the predicate together with all reductions necessary to match the antecedents of the warrant to the facts of the case [p. 43].”

Now, what does this have to do with the jurisprudential concept of *ratio decidendi*? In traditional jurisprudence, identifying the *ratio* of a case was a way to specify which components of a precedent should be taken as authoritative in subsequent cases. Although the literature on this subject was immense, Karl identified four characteristics of the *ratio* that most legal philosophers would agree upon: (1) it should include the propositions of law that are necessary for the decision, as opposed to mere *dicta*, which are propositions of law that could be negated without changing the result; (2) it only rarely consists of a single proposition, but usually includes a range of propositions, at various levels of abstraction; (3) it should be "grounded in the specific facts of the case [p. 41]"; and (4) it should include "not only the precedent's material facts and decision, but also the theory under which the material facts lead to the decision [p. 42]." Karl then argues that the entire reduction-graph, and all of its components, i.e., the entire SLD-tree that constitutes the "justification" of the legal decision, should be taken as the *ratio decidendi* of the case.

Actually, the fourth characteristic in the preceding list was, at one time, somewhat controversial. Arthur Goodhart had argued that the *ratio decidendi* should include only the material facts and the outcome of a case, and not the reasons given by the judge for his decision [98]. Others, such as Rupert Cross [83], had argued that the theory under which a case was decided was equally important in its role as a precedent in subsequent cases. Using the English case of *Bourhill v Young, A.C. 92 (1943),* as an example, Cross showed that there were two theories under which that case could have been decided, one in which liability was denied because the defendant did not owe a duty of reasonable care to the plaintiff, and one in which liability was denied because the defendant's conduct was not a proximate cause of the plaintiff's harm, and it would make a difference in future hypothetical cases which of these two theories was employed. In Figure 3 and Figure 4 of his paper, Karl illustrates these two theories of *Bourhill v Young* using his reduction-graph model, and shows that the complete SLD-trees are needed to capture Cross's distinction, since the material facts of *Bourhill v Young* are exactly the same in both trees.

In the final section of the paper, Karl takes this argument one step further, and applies it to several contemporary examples of case-based legal reasoning. His main point here is that a "pure exemplar-based" theory of precedent would have to be consistent with the Goodhart view of *ratio decidendi*, and therefore subject to the critique of Cross in a case such as *Bourhill v Young*. Four approaches to exemplar-based reasoning are considered: (i) *structural similarity* [134]; (2) *dimensional analysis* [16]; (3) the *nearest neighbor* classification rule [172,266]; and (4) my own *prototype-plus-deformation* model [178]. For each approach, Karl argues that the authors of these studies have adopted the Goodhart view, implicitly and unavoidably, thus making it impossible to represent the distinction that Cross wants to make in *Bourhill v Young*.

Karl Branting published another article[6] on his reduction-graph model in a special issue of the journal *Artificial Intelligence* in 2003 [69], but the extensive jurisprudential debate was not included. There is a short discussion of *ratio decidendi*, one citation in the text to Rupert Cross, and an illustration of the alternative theories in *Bourhill v.Young*, but there is no mention of Arthur Goodhart, and no evaluation of the reduction-graph model in terms of the Goodhart/Cross debate. Curiously, there is not even a citation to Karl's previous ICAIL 1993 paper! Perhaps Karl had decided by then that the Goodhart/Cross debate was ancient history in jurisprudence, and no longer relevant. He remarks that the concept of *ratio decidendi* was part of "the *orthodox* view of precedent," and notes that "many legal scholars would argue that the orthodox view is a drastic simplification of the actual use of precedents in legal discourse and problem solving." But even if our jurisprudential theories are much more sophisticated today, our AI models are simple enough that they can benefit from the criticism of previous generations of legal scholars.

[6] An expanded version of the ICAIL 1993 paper was also published in *Artficial Intelligence and Law* [68]

Evaluating a computational model using the standards of a jurisprudential theory sets a good precedent, in my opinion, for future research.

5.3 Edwina L. Rissland, David B. Skalak and M. Timur Friedman. BankXX: Supporting legal arguments through heuristic retrieval [231]. *Commentary by Trevor Bench-Capon*

One of the most remarkable things about BankXX [231], but see [234] for a fuller description, is that it did not spawn a host of imitators. Unlike other related case-based systems such as HYPO, CABARET and CATO, which have inspired a great deal of work building on them and reconstructing them, and attempting to capture their logic, BankXX seems to have inspired little further development and few if any imitators. None the less BankXX is an important addition to our understanding of case-based reasoning and argumentation built on it.

Whereas CABARET ([228] and see 3.4), for example, operates top down BankXX offers an experiment in *bottom up* argument creation. BankXX takes a highly interconnected network of legal information and crawls through it using best first heuristic search to collect what it needs to construct an argument. The program is based on the general knowledge required to answer three questions:

- What domain knowledge is available?
- What basic pieces are needed to form an argument?
- What makes an argument a good argument?

The knowledge is in the domain of individual bankruptcy law, and includes cases, represented as factual situations, as bundles of citations, as scripts giving stereotypical stories, as sets of legal factors and as measures of their prototypicality. Each of these aspects occupies its own *space* and the information is linked both within and across spaces. Legal theories are also represented in terms of the legal factors they use. In [231] there were twelve argument pieces to act as the building blocks. These included cases decided for and against the current viewpoint, and various measures of match, applicable legal theories and prototypical stories. Finally arguments were assessed according to argument factors, of which eight were implemented in [231], including the win record of the theory used, the strength of the citations and the strength of the best case analogies. Three linear evaluation functions were used to guide the search. At the domain level the function determined how well a node contributed to information known to be useful to an argument. At the argument piece level, the function captured whether the particular knowledge in the node could be used to complete a component of the argument not yet filled. At the argument level the function tested whether including the piece would improve the quality of the argument according to the argument factors. The output from BankXX was a collection of argument pieces. This did involve some expert interpretation to see what the argument might be: it is possible that some post processing to put them into an English template might have widened its appeal.

The conclusion of [231] was that although the authors expected that some combination of the top down approach and this bottom up approach would be necessary for creating arguments, BankXX did identify useful sets of components and comparisons with the argument pieces used in actual opinions were encouraging, and that further exploration was needed. Unfortunately, as we have noted, there was very little follow up. Subsequent AI and Law work which generates arguments is almost all top down, using an argumentation scheme to determine the pieces needed, and a knowledge based system to instantiate these templates. An early example is [36], in which a set of argument relations are extracted from a knowledge base and then organised into an argument using Toulmin's schema, but a number of other systems along these lines have been developed using a variety of argumentation schemes. All of them, however, use information organised as a rule base, rather than searching a collection of information to harvest potentially useful information. Heuristic search has also been little used: Karl Branting's GREBE [66] had previously used A*, but on a very different search space (mappings from the problem case to a representation of precedents structured as a semantic net. Alison Chorley's AGATHA [78] also uses A* to construct an argument, but there the argument takes the form of an adversarial dialogue and the search space is the game tree.

The time does, however, appear ripe to revisit BankXX. The combination of arguments and stories we find in BankXX has appeared again in the work of Bex and his colleagues (e.g. [47] and see 14.3). Argumentation schemes abound, and finding suitable pieces to instantiate them is an obvious way to use them. Most importantly the Internet, which was very much in its infancy in 1993 (Google was not founded until 1998, and even Altavista was not founded until 1995) has come on in leaps and bounds since then. So now we have a ready made collection of legal information organised as a highly connected network out there just waiting to be harvested. Certainly the three questions used to drive BankXX and perhaps its evaluation functions would make an excellent starting point for doing that.

5.4 Thomas F. Gordon. The Pleadings Game; An Artificial Intelligence Model of Procedural Justice [101]. *Commentary by Henry Prakken*

By 1993 the idea of using nonmonotonic logics as a tool for formalising legal argument was already somewhat established. In [100][7] Tom Gordon added a new topic to the research agenda of the formalists in AI and Law: formalising the procedural context of legal argument. Gordon attempted to formalise a set of procedural norms for civil pleading by a combination of a nonmonotonic logic and a formal dialogue game for argumentation. The resulting Pleadings Game was not meant to formalise an existing legal procedure but to give a "normative model of pleading, founded on first principles", derived from Robert Alexy's [8] discourse theory of legal argumentation.

The Pleadings Game had several sources of inspiration. Formally it was inspired by formal dialogue games for monotonic argumentation of e.g. Mackenzie [173] and philosophically by the ideas of procedural justice and procedural rationality as expressed in e.g. [8], [223] and [265]. For example, in [265] Toulmin claimed that outside mathematics the validity of an argument does not depend on its syntactic form but on whether it can be defended in a rational dispute. The task for logicians is then to find procedural rules for rational dispute and they can find such rules by drawing analogies to legal procedures [265, p.7].

In AI Ron Loui had in 1992 started circulating his Process and Policy paper (finally published in 1998 as [165]), in which he argues that Rescher's and Toulmin's views should be taken seriously by nonmonotonic logicians: the 'logical' layer of nonmonotonic logics should be embedded in procedures for rational dispute and nonmonotonic logicians should investigate conditions under which these procedures are "fair and effective". Crucially, such procedures should unlike nonmonotonic logics allow for the non-deterministic construction of a 'theory' during a dispute. With the Pleadings Game, Gordon presented the first thorough formalisation of these ideas in AI & Law. There had been one earlier proposal by [122] but this attempt was still sketchy. It was later more fully developed in [124, 162].

[7] This paper was a summary of Gordon's PhD thesis of the same year, which later appeared in revised form as [102].

Besides a theoretical goal, Gordon also had the aim to lay the formal foundations for a new kind of advanced IT application for lawyers, namely, mediation systems, which support discussions about alternative theories by making sure that the rules of procedure are obeyed and by keeping track of the arguments exchanged and theories constructed.

At a time when I was fully devoted to the application of nonmonotonic logics to law, Gordon's paper was one which broadened my mind: it made me start reading and thinking also about formalising procedural games. This did not immediately lead to publications. Although I found the literature on this topic fascinating, I was somewhat confused by the variety of the proposed systems, both in AI and Law and in other areas, such as multi-agent systems: unlike logical models of a argumentation (where [86] was a strong unifying force), the various dialogue systems were rather ad hoc and did not have much in common. This was one of the reasons that I did not publish on this topic until my [203] in 2001. My aim to discover underlying principles of dialogue games for argumentation eventually led to my [206] in 2005, which I then adapted for legal procedures in [207] in 2008. Although I think I have made some progress, the state of the art is still less advanced in dialogue games for argumentation than in logical models of argumentation.

While Gordon's work was initially followed by a considerable amount of other work (besides the above-cited work also by e.g. [29]), the topic has now in AI and Law somewhat gone out of fashion (in contrast to the field of multi-agent systems, where dialogue games for argumentation are intensively studied as a way to regulate and promote rational interaction between intelligent artificial agents). In my opinion this is unfortunate, since both Gordon's theoretical and practical aims are still important today, as I experience in my own current work with others on legal reasoning about evidence. Clearly legal evidential reasoning is defeasible, so rational models of legal proof should have some logical account of defeasible reasoning as a component. However, miscarriages of justice in legal proof are often not caused by inferential errors but by investigative flaws or by the failure to ask the right questions at trial. A well-known problem is that of confirmation bias: if the police thinks they have caught the right suspect, they often only look for evidence that confirms their suspicion, but if one does not look for falsifying evidence one will not find it. This could mean that although logically (given the created information state) the suspect might be proved guilty, rationally the decision to convict is still flawed. A full theoretical account of rational legal proof should therefore also address procedural and investigative aspects. One modest attempt to formulate a procedure for discussing crime scenarios is [53] but much more work should be done.

Moreover, the idea to develop mediation systems for crime investigations and criminal trials is very promising (likewise [155] for legal procedures in general). Police forces frequently use software for drawing time lines and relational schemes and they feel the need for extensions of such software with means to connect their analyses with evidential arguments. A demonstrator prototype of a system with such functionality was developed by Susan van den Braak [64]. In my opinion the state-of-the art is now mature enough to develop practically useful systems of this kind, both for crime investigators and for lawyers.

5.5 Ronald P. Loui, Jeff Norman, Jon Olson, and Andrew Merrill. A Design for Reasoning with Policies, Precedents, and Rationales [166]. *Commentary by Floris Bex*

In [166] Loui and his colleagues propose a general formal model of argument and explore the suitability of this model for reasoning with policies, precedents and rationales. While the contributions and conclusions of the paper itself were fairly modest at the time, the ideas presented were, in my opinion, the spark that ignited some of the more interesting and unifying research in AI and Law. Furthermore, these ideas can also be seen to have an impact on the field of argumentation and commonsense reasoning in general.

Research on legal Case-based Reasoning (CBR) has been one of the mainstays of AI and Law since the late 1980s. This research and the applications based on it, such as CABARET [230], HYPO ([16]) and CATO [5], focuses on the dialectical process of citing and comparing legal cases, and on heuristics for case-based reasoning. As such, it is clearly grounded in a common law framework in which judge-made law in the form of precedent cases is the norm.

A second main strand of the AI and Law research in the past twenty years has been the development of rule-based approaches to defeasible reasoning in the law, the main proponents of which have been Gordon [102], Prakken and Sartor [209] and Hage [118]. This work also models a dialectical process, but instead of reasoning with legal cases it focuses on legal rules and how these rules can be incorporated in a logical framework for reasoning with incomplete or inconsistent knowledge. In this sense, the ideas expressed in this line of research clearly stem from a continental legal perspective, in which rules devised by a legislator guide the process of legal reasoning.

The paper by Loui and his colleagues came at a point when legal CBR had already started to take off as a serious area of research (HYPO dates from the late eighties and CABARET from the early nineties) whilst the influential rule-based approaches were still in their infancy (Prakken's and Gordon's dissertations were published in 1993). Loui and colleagues can therefore be considered pioneers in trying to unify models of defeasible argument with the work on CBR; while Ashley, Rissland and others directly referred to the concept of argument they did not, as Loui and colleagues did, give a full logical account of it in the style of formal frameworks for defeasible reasoning (e.g. [198]). Loui and colleagues thus paved the way for later research in which case-based reasoning is modelled as a type of defeasible argumentation ([167], [212] and [41]).

A second important contribution of the 1993 paper by Loui and colleagues is that they for the first time formally model what they later in [167] call compilation rationales: rationales for rules that compile pieces of past disputations, past lines of argument, past preferences of one argument over another or past projections from cases. For example, the defeasible rules

$vehicle(x) \rightarrow privateTransport(x)$ and
$vehicle(x) \wedge privateTransport(x) \rightarrow \neg allowedInPark(x)$

allow for a two-step argument with conclusion

$\neg allowedInPark(t)$,

where t is some instantiation of x. Loui and Norman's compression rationale states that the two rules may be compressed into one defeasible rule, namely

$vehicle(x) \rightarrow \neg allowedInPark(x)$.

These two contributions - incorporating CBR into a logical rule-based argumentation framework and defining rationales that state how rules may be compressed or specialised - have relevance not just for AI models of legal reasoning but also for more general models of commonsense reasoning. Consider the idea of commonsense knowledge about the world that can be used to form opinions, decide on actions and generally reason about the world and our place in it. Now, this commonsense knowledge has in the past variously been modelled as scripts or stories ([247]) and as rules, defaults or generalizations ([198]; [222]). For example, the well-known restaurant script details the events and agents involved in a typical restaurant visit, thus providing a 'holistic' structure that captures a piece of world knowledge. Generalizations also capture world knowledge but are typically more 'atomistic', for example, rules such as 'if you don't pay in a restaurant the owner will usually call the police'. As Bex, Bench-Capon and Verheij have argued recently ([55]), scripts and stories are essentially factual, non-legal cases in the vein of [5], and the same argument moves available in CATO can be applied to cite and distinguish these factual stories. Furthermore, [49] showed that Loui and Norman's compression rationales are not just relevant for legal rules but also for commonsense rules (i.e. generalizations). For example, the rule captured by 'if an eyewitness testifies P then usually P is the case' can be decompressed into three separate rules that capture the veracity of a witness (if a witness says 'I remember I saw P then P), her objectivity (if a witness remembers he saw P then P) and her observational sensitivity (if a witness saw P then P). Thus, if we extend the work by Loui and colleagues to apply to not just legal cases and rules but also to non-legal cases and rules, we have a flexible and powerful framework that encapsulates different ways of reasoning with commonsense knowledge as defeasible argumentation.

5.6 Trevor Bench-Capon. Neural Networks and Open Texture [28]. *Commentary by Bart Verheij*

Is it possible to learn the rationale underlying decisions in an open textured domain, such as the law, given only a set of decided cases? It is this important question that Bench-Capon (1993) investigates in [28]. He investigates the question using neural networks. At the time, neural networks were a popular research topic, as they had helped solve problems, e.g., recognition of hand-written characters, that seemed unsolvable using logic-based artificial intelligence. The paper is a beautifully written, self-contained essay, and contains lessons about the automatic learning of rules from cases that are still valuable.

The learning experiments are performed on cases concerning a welfare benefit to be paid to senior citizens visiting their hospitalized spouse. Past decisions are artificially generated (by a LISP program), constrained by six conditions. The setting is fictional, in the sense that it does not reflect an actual welfare policy, but the example has been carefully designed in order to be able to investigate different kinds of legal conditions. For instance, there is the Boolean condition that the two persons involved should be married. There is a threshold condition, namely that the couple's capital resources should not exceed some fixed amount. There are also dependencies between variables, such as between age and sex: the person should be of pensionable age, 60 for women, 65 for men.

Bench-Capon investigates two questions: Can the neural network be trained to decide cases on the basis of a given set of decided precedent cases? And can the decisions proposed be justified in terms of the conditional constraints used to generate the cases?[8]

Bench-Capon answers the *first question* about learning correct decisions with a resounding 'Yes'. Quoting the paper:

> Neural networks are capable of producing a high degree of success in classifying cases in domains where the factors involved in the classification are unknown. (p. 296)

[8] Bench-Capon discusses a third question, namely whether we can derive rules from the networks. Because of the way in which he addresses this question, it is not further discussed here.

He reports success rates of around 99% for networks with one, two or three hidden layers, with networks trained on 2400 generated cases and tested on 2000 cases. He shows that in his setup irrelevant factors do not strongly reduce performance.

The *second question*, Can the network's decisions be justified?, is answered with interesting nuances. Bench-Capon's original training set shows that for the one and two hidden layer network designs *sex and age do not matter* for the decision: when the other constraints are satisfied, the network proposes the decision to pay the benefit, for almost all ages, even for ages close to 1. In the three hidden layer design, there is a relevant correlation between sex, age and the decision proposed, but the age difference associated with the dependency is 15, not 5. Also everyone above 40 is paid the benefit, instead of everyone above 65.

In order to explain this finding, Bench-Capon analyzes his material and finds that already four out of six factors can explain about 99% of the cases. Bench-Capon continues his experiments on the basis of sets of generated cases with a more careful distribution over the satisfaction or non-satisfaction of the different conditions. He finds that he can hereby steer the network's performance to more closely reflect the sex-age constraint as it was actually used to generate the cases.[9]

He draws an insightful and important conclusion. That the two features were missed 'can only be detected if we have some prior knowledge of the domain which allows us to say this: otherwise we have no way of telling that the four conditions that were discovered were not in fact the whole story' (p. 294–295). This is almost like saying that the rationale shown by the neural networks can only be estimated by knowing about the rationale beforehand. Put yet another way: rules mimicking past decisions need not determine the rules used to make the decisions.

An interesting, and still relevant, discussion of Bench-Capon's paper can be found in the 1999 special issue of the journal *Artificial Intelligence and Law* [196], devoted to neural networks (and fuzzy reasoning). In a position paper on the contemporary state of the art of applying neural networks to the law, Hunter [137] claims that neural network applications to the law were in general flawed, in particular, because they were often based on inappropriate data. In contrast, Bench-Capon's paper is praised as being 'methodologically most appropriate', but, says Hunter, it is just 'about the use of neural networks to simulate necessary conditions, since the training set and the verification set were simply drawn from rules'. From this, Hunter concludes that Bench-Capon is not about *legal* neural networks at all, for they only model an already given rule-encoded doctrine — or, when learning fails, not even that.

I think that Hunter misses the importance of that part of Bench-Capon's results, and of their relevance for law. Still, Hunter is right that some typically legal phenomena are not covered. Indeed, Bench-Capon's networks show no development in the light of new circumstances, cannot handle new insights, cannot incorporate landmark precedents, do not incorporate values that steer decisions. Interestingly, each of these themes is particularly addressed by one of Bench-Capon's other core research topics: case-based argumentation (with Ashley's [16] a milestone).

I wonder what Bench-Capon's reaction at the time was when reading Hunter's paper. I believe that in his heart he agreed; his neural networks were not sufficiently 'legal'. For as we all know, in the years that followed, Bench-Capon devoted much of his time to what was

[9] A question remains however: the network has 45 and 50 as significant ages, instead of 60 and 65. Bench-Capon gives no explanation for this oddity: is it a systematic consequence of the learning rule used, perhaps a small bug in the set up?

missing, and developed dynamic case-based reasoning techniques, with the incorporation of values a key innovation (see, e.g., his influential [31]).

At the same time, Bench-Capon (and the field) remains interested in the learning of rules underlying legal decisions, a recent contribution being the PADUA system [278]. In this work, the same legal conditions are used as in the 1993 paper, this time using training sets that can contain errors.

Today Bench-Capon's 1993 conclusion seems to stand strong: The patterns in a given set of decisions do not determine the rules that led to the decisions. Perhaps the time has come to reconsider the dynamical relations between the logic of decision-making and the probabilities associated with data description?

5.7 Giovanni Sartor. A Simple Computational Model for Nonmonotonic and Adversarial Legal Reasoning [243]. *Commentary by Guido Governatori*

In [243] Giovanni Sartor addresses the issue of the representation of non-monotonicity in legal reasoning. He does so by providing a computationally oriented model in PROLOG. The nineteen-nineties can be considered the golden age for non-monotonic reasoning. A large amount of research in the general field of artificial intelligence was dedicated to this then emerging and promising area. Artificial Intelligence and Law was not immune from this trend and scholars like Sergot, Prakken, Gordon, Bench-Capon (and many others) were interested in exploiting these developments. A contribution of [243] was an analysis of areas of law and legal reasoning requiring non-monotonic reasoning. A key observation was that resolution of conflicts in non-monotonic reasoning is based on the concept of preference ordering over elements in conflict, and this is no different from the resolution of what D. Ross called *prima facie duties* when they conflict, i.e., conflicting norms. However, normative systems are dynamic, there can be multiple concurrent legal sources, and legal languages often leave space for semantic indeterminacy (i.e., multiple interpretations are possible). These considerations led Giovanni Sartor to state that formalisations of legal reasoning need inference procedures taking into account an ordering relation, and that ordering relation should be obtained using many criteria that have to be "harmonised". The paper proposes a solution for this issue.

The technical solution developed in [243] is that norms are represented by rules with the form $n : p_0 \leftarrow p_1 \wedge \ldots \wedge p_n$, where n is the name of the rule, where $n = r(X_1, \ldots, X_m)$, r is a new function symbol and $X_1, \ldots, X_m$ are the free variables appearing in the rule. The technical device just presented allows us to represent selective exceptions directly in the object language. A rule $r(X_1, \ldots, X_n) : p \leftarrow q$, is then translated into the pair $r(X_1, \ldots, X_n) : p \leftarrow q \wedge \mathit{applicable}(r(X_1, \ldots, X_n))$ and $r(X_1, \ldots, X_n) : \mathit{applicable}(r(X_1, \ldots, X_n))$ where the fact $\mathit{applicable}(r(X_1, \ldots, X_n))$ asserts that all instances of the rule $r(X_1, \ldots, X_n)$ are applicable. The introduction of this predicate offers a flexible and powerful feature. We can use it to specify conditions under which a rule can or cannot be used. The second advantage of the introduction of rule names is that they can be used in predicates expressing preferences over the ordering of norms, and that it is possible to have rules whose head is a preference predicate.

Preferences over rules have been a standard feature in non-monotonic reasoning. The novelty of the paper is that preference handling is dynamic, and we can reason about the preference rules. With a few exceptions, e.g., [210] and [13], the idea has been neglected until recently, when some impulse in this direction has arisen from work on extended abstract argumentation [183] and work on revision of preferences [108].

6 University of Maryland 1995

The Fifth ICAIL was held at the University of Maryland (inside the beltway of Washington D.C) from May 21st to 24th, in the memorably named Volunteer Fire Fighters Room. Thorne McCarty was Conference Chair and Trevor Bench-Capon was Programme Chair. The programme committee eventually comprised eleven members, five US, five Europeans, representating five different countries, and a Canadian. An Australian member was invited but unfortunately was in the end unable to participate. Invited speakers were the founder of Logic programming, Bob Kowalski, and (arguably) the founders of AI and Law, Bruce Buchanan and Thomas Headrick. The panel took place at an evening reception at the Mayflower Hotel in Washington itself, and addressed the future of legal information systems. This conference also firmly established the importance of the Netherlands in the AI and Law community, with nine papers (over a quarter) coming from various groups in that relatively small country. Eight papers have been chosen from this conference, ranging from information retrieval (6.7) to meta-inference (6.1) and an early approach to ontologies (6.3). Argumentation remained important (6.2, 6.4 and 6.6) as did the need to detect and cope with change in the law (6.5).

6.1 Haijme Yoshino. The Systematization of Legal Meta-Inference [287]. *Commentary by Michał Araszkiewicz*

In this paper [287] Professor Hajime Yoshino presents his theory of legal meta-inference and argues for the thesis that if what is referred to as legal meta-knowledge is described exactly, it is unnecessary to apply any inference system to legal reasoning other than classical first-order logic. In this respect, the paper's methodology is rooted in earlier research in AI and Law and especially in the work of Yoshino himself (and in particular, his series of legal expert systems called Legal Expert System (LES), e.g. [286]). Because Yoshino focuses on representation of legal knowledge by means of rules and employs Prolog implementation, his work is cognate to classical 1980s and early 1990s contributions to the theme of logical representation of legislation (most famously, [255]). Consequently, Yoshino's proposal is based on a skeptical view of the use of of nonmonotonic logic or default or defeasible logic, which became influential as a field of research within AI and Law in early nineties (e.g. [199] and [243]). The method and flavor of Yoshino's contribution is rooted in his broader legal-philosophical project which he refers to as Logical Jurisprudence. The theory is inspired to some extent by the work of Hans Kelsen, but it should be emphasized it is an original contribution which differs considerably in many respects from Kelsen's *Pure Theory of Law*.

This paper is based on the fundamental concept of legal meta-inference which controls legal inference. This concept is used as a tool which can be useful in construction of a legal knowledge base. Legal knowledge increases over time and it is subject to change: at different times we should draw different conclusions. This suggests that legal inference has a nonmonotonic character, but Yoshino defends a thesis that it is possible to represent all legal inference by means of first order classical logic, given that legal meta-inference is adequately represented in legal knowledge system. Legal meta-inference systematizes and guides legal inference. Yoshino adopts and modifies the jurisprudential distinction between legal object rules (binding on the addressees, such as citizens) and legal meta-rules, which operate on other legal rules, in particular prescribing their validity. The concept of legal validity plays a pivotal role in Yoshino's theory. He admits that in earlier version of his conception he used more dedicated predicates about rules (such as 'applicable' and 'formally applicable') but he abandoned this idea because the concept of legal validity suffices to attain all the theory's aims. Yoshino claims that the predicate 'is valid' in the domain of rule-based reasoning plays a similar role to the concept of truth value in logic. According to the paper, legal meta-rules prescribe (1) the scope of validity of legal rules (in terms of 'time', 'space', 'person' and 'matter') and (2) the priority relations between different types of rules. Formulation of different types of meta-rules leads Yoshino to the most general rules which regulate the issue of legal validity in terms of 'becomes valid' and 'becomes null' predicates. According to [287], these abstract principles are exhaustive (no other rule of this level of generality is needed to deal with the problem of validity of legal rules) and may be compared to Kelsen's concept of the basic norm (*Grundnorm*). Importantly, these rules are not legal rules, but they are presupposed by lawyers and they make legal meta-inference possible. Undoubtedly, the formulation of these abstract principles is an important contribution of the paper. Yoshino employs the so-called Compound Predicate Formulae (CPF) to represent legal knowledge (more on CPF in [288]). Legal reasoning is performed by a legal meta-inference engine. The inference engine checks the validity of every rule used in the reasoning (that is, whether this rule is derogated or not) and yields, on the basis of valid rules, an answer to the legal question under consideration. The paper presents interesting examples of representation of legal reasoning which confirms the usefulness of legal meta-inference.

The paper is an example of dealing with complicated problems of rule-based reasoning by means of classical first-order logic. Given the historical context and the development of alternative logical tools for modeling legal reasoning, including for instance defeasible logic, dialogue logics, argumentation frameworks and so on the paper is an example of attaining important results concerning representation of legal knowledge based on minimal set of concepts and assumptions. The line of research presented in the paper was further developed by Yoshino, resulting in many revisions and developments of the presented framework (for a recent account see [289]).

The contribution described here is an interesting, relatively early account of legal meta-reasoning and Yoshino was one of the first scholars in the AI and Law community who emphasized so strongly the importance of this issue. Nowadays, meta-argumentation and legal reasoning concerning validity is an object of interest of many researchers, albeit in different formal settings. For example see, for meta-argumentation, [182] and, for temporal legal reasoning, [188].

6.2 Henry Prakken. From Logic to Dialectics in Legal Argument. [200]. *Commentary by Trevor Bench-Capon*

In the eighties and early nineties there were several disputes in AI and Law which related to the appropriateness of using logical tools in AI and Law. Proponents of case-based reasoning would argue that logic neglected the adversarial nature of legal reasoning, and that it was unable to handle indeterminacy and conflict (e.g. [45]). Others claimed that although logic might be useful it would have to be a nonmontonic logic because, for example, law was structured around exceptions to general rules (e.g. [99]). Thirdly there were those who argued that emphasis should be placed on the procedural nature of legal decision making (e.g. [101] and see 5.4). In [200] Henry Prakken showed that all these disputes are unnecessary and so paved the way for the highly fruitful use of logical tools to address issues relating to adversarial and procedural aspects of legal reasoning and to issues of indeterminacy and conflict that have since been an essential feature of AI and Law.

The main point of Henry's paper is to distinguish three levels required for adversarial legal reasoning.

1. The *logic* level, which generates arguments;
2. The *argument* level, which organises these arguments and identifies attack relations between them; this determines the acceptability of arguments at a given point in the debate;
3. The *dialogical* level, which determines how the arguments and attacks from the level below can be deployed in a dispute. This level moves the debate forward and refers to the level below to determine the acceptability of arguments at the current stage.

Once we have separated these three levels we can see more clearly what should be done at each of them. The logic level generates arguments, and it is quite possible that some of these arguments will conflict: perhaps because we have exceptions, or because interpretations differ. But it is not the business of this level to reconcile such conflicts. The arguments remain arguments, even if we will later reckon them to be defeated. The conclusion of [200] is that the logic not only can be monotonic, but should be monotonic: any nonmonotonic behaviour required comes from the higher levels.

At the argument level these conflicts are recognised, and an attack relation between arguments formed, to give rise to an Argumentation Framework as introduced by Phan Ming Dung [86][10]. In [200] we find two types of attack: *rebuttal*, where two arguments license contrary conclusions, and *undercut* where the argument attempts to show that the attacked argument is invalid (because, for example, a rule used is inapplicable). The former kind of attack is symmetric and the latter asymmetric. Both these kinds of attack continue to this day to be recognised in abstract argumentation systems, along with a third kind, *premise attack*, where the attacking argument concludes the contrary of a premise of the attacked argument: such arguments would be seen in [200] as attacking a subargument, either by rebuttal or undercut. The status of arguments can be determined from this framework and in [200] the acceptability of arguments was determined at this level.

Finally the procedural level tells us how to conduct a dispute according to the rules of the legal process being considered, and so dynamically constructs the theory from which arguments are generated. Arguments may come to be, and cease to be, acceptable as changes to the theory will change the arguments determined as acceptable at the second level. For example if the procedure introduces an attacker of a currently acceptable argument, this argument will cease to be acceptable, but can be reinstated when the procedural level adds the means to generate an attacker of this attacker. In this way nonmonotonic behaviour is achieved by additional arguments and attacks extending the argumentation framework, and so changing the status of arguments within it.

Abstract argumentation as expounded in [86] can determine the status of arguments according to a variety of semantics, such as grounded, sceptical preferred and credulous preferred. In the years since 1995 it has been found helpful to devise procedures in the form of dialogue games which determine the status of arguments according to various semantics. For example [271] gives dialogue games for credulous and sceptical preferred semantics. We might therefore be tempted to move the determination of argument status to the procedural level so that this level can be used to determine and explain which arguments from the argument level should be accepted, perhaps also incorporating elements of legal procedure into these games. For example, a legal-credulous game might be used for determining the right to appeal, and a legal-sceptical game for deciding whether the appeal should be successful. More in keeping with the spirit of Henry's original paper, however, would be to

[10] In fact this journal paper was still in press and so [200] refers to Dung's 1993 IJCAI paper.

divide the argumentation level so the notions of attack and, where preferences can be used to defend arguments as in [31], defeat are separated from the semantics chosen to determine acceptability while retaining a separate level for the legal procedures governing the dynamic aspects of the dispute.

Once we start to think in terms of these different levels, we can see that many of the controversies that existed at the time arose from a conflation of things which are properly kept separate. For example we should not expect a logic to find arguments *and* identify conflicts *and* resolve these conflicts, as nonmonomtonic logics are supposed to do. Trying to do this is both wrong and futile. Wrong since it will deny that some arguments are arguments at all: if the logic also determines acceptability there are never two sides to the question whereas in law this should be the normal situation. Futile since the search for a generally applicable conflict resolution principle is misguided. When Lex Specialis and Lex Posterior conflict, or when open textured predicates need to be resolved, as [200] says:

> if for solving such conflicts any guidelines are available at all, they are of such a diverse and tentative nature that there is ample room for debate

and we should not wish to curtail such debate by decisions built into in our logic.

Similarly we want our procedures out there in the open, so that we can see what the effects of different procedures would be and explicitly debate questions such as who has the burden of proof, or the standard of proof required to discharge such burdens. Having the distinct levels levels opens the door to a plurality of views, and enables argument about these views, at the appropriate level, so when we make choices we are aware that we are making a choice, and have the opportunity to justify it.

Nowadays this division is more or less accepted into the culture, and people rarely speak explicitly of these levels, but simply proceed through them. For example [24] generates arguments by instantiating an argument scheme against a model (logic level), organises them into a (value-based) Argumentation Framework, next evaluates their status using preferred semantics and then offers the resulting preferred choice for public critique. Only when someone conflates the levels do we need to remind them of the need to keep them separate. We might want to further divide some levels: for example we might want to divide the argument level using a choice of audience to move from a value-based argumentation framework based on an attack relation to an audience specific framework with only a defeat relation (see [31]), or to allow for different games to apply different semantics as discussed above. Or we might want to divide the dialogical level to distinguish the rules of the game (playing the game) from strategic rules (playing the game *well*). More levels, properly motivated can be an advantage because they enable finer distinctions which may be important: when we say *you cannot move your queen there*, we might mean either that the move is illegal or that it will lose the game and we have to recognise that playing to lose is possible.

The paper also suggested that non-deductive arguments, such as those based analogy and induction, should be treated as a source of additional premises, but we will not discuss this suggestion here. Instead we have concentrated on the important distinctions that provided a means to put an end to some disputes based on the confusion of what should be distinct aspects of legal reasoning. In so doing [200] laid the foundations for the logical analyses of case-based reasoning that have proved so useful since: one may mention as examples [212], [41] and, most recently, [135].

6.3 Andre Valente and Joost Breuker. ON-LINE: An Architecture for Modelling Legal Information [267] *Commentary by Enrico Francesconi*

The study of the relationships between intelligence and knowledge represents a topic widely discussed in the literature of several disciplines (from philosophy, psychology, cognitive sciences, biology, to applied sciences as robotics, computer science and artificial intelligence).

Without delving into the philosophical implications which the distinction of the specific characters of one with respect to the other would bring about, in computer science it is a common assumption that intelligence requires knowledge [224]. This means that knowledge modelling is an essential pre-condition to develop intelligent systems. The attempt to formalize knowledge so as to make it amenable for computation finds in specific domains, including the legal domain, a privileged environment of application, since they are characterized by a specialized and technical kind of information. Nevertheless the effective implementation of intelligent legal information systems still represents a significant challenge.

In the last 25 years of AI and Law research, knowledge modelling has been widely addressed in the literature, starting from rule-based systems and developing into ontology-based legal information services and semantic web applications.

From this point of view, in the mid ’90s, Valente and Breuker presented ON-LINE [267] which can be considered one of the first attempts to implement a legal information management system whose architecture included legal information storage and retrieval facilities, as well as legal reasoning functions. The authors emphasised that the key of such an integrated service was given by the theoretical framework provided by a legal ontology expressing legal norms. With such an architecture, they acknowledged the key role of legal knowledge modelling in managing legal information for legal assessment tasks, as a step forward with respect to rule-based representation systems. The authors thereof shifted the problem of legal assessment from a rules-oriented formalism, able to represent legal consequences about the application of relevant norms to a case, to a knowledge representation based on a specific theory of ‘what law is made of’.

This was the ground for the elaboration of a Functional Ontology of Law which distinguishes a number of primitive types of legal knowledge, independent from the formalism used for reasoning with it and organized according to their functional characterization (normative, domain oriented, meta-level knowledge, etc.). The Functional Ontology of Law was conceived as an artifact of abstract legal concepts, providing a foundational ontology for the legal domain, amenable for reuse and sharing.

This knowledge modelling was applied to provide a formalized, language independent, representation of legal information, able to distinguish between textual knowledge and conceptual knowledge, as well as the links between them. Such a distinction represents the basis for resolving ambiguities, multiple interpretations of the law, as well as the representation of legal documents (precedents, jurisprudence, commentaries, cases, etc.) which improves usability and re-usability of legal information, combined with the development of advanced services of legal analysis.

In particular the anchorage of textual legal information, stored as atomic objects as paragraphs, articles, etc, with conceptual legal knowledge represented in ontologies, can be considered as an *ante-litteram* Semantic Web oriented application in the legal domain, paving the way for a fertile research trend. We will see this trend continued in Section 7.1.

6.4 Arthur M. Farley and Kathleen Freeman. Burden of Proof in Legal Argumentation [88]. *Commentary by Thomas F. Gordon*

Farley and Freeman's paper on "Burden of Proof in Legal Argumentation" [88] has proven to be highly influential because of its computational model of several legal proof standards, including scintilla of evidence, preponderance of the evidence, and beyond reasonable doubt. This was, to my knowledge, the first computational model of proof standards and inspired, as we will see, a line of research on this topic.

Proof standards, however, were only one part of the work, which aimed to be a more comprehensive computational model of the structure and process of argumentation, covering:

- the invention of arguments from propositional rules, called "warrants", following [265]
- the construction of argument graphs, with several kinds of attack relations between arguments
- the evaluation of arguments, using argument weights and the model of proof standards
- and modeling the dialectical process of argumentation, as a two-party game.

Farley and Freeman's work had much in common with some contemporaneous work on models of argumentation in the field of AI and Law, including [34], [100], [166], [123], [163], [154], and [209]. It is not possible in these few pages to compare these works in detail, or to try to understand just how they influenced each other. To a considerable extent they appear to be aware of and cite each other's work. Farley and Freeman's work began with an interest in legal reasoning, but was inspired and informed mainly by work on argumentation in philosophy, in particular by [265], [223] and [198], along with work on nonmonotonic logic, and is intended to be broadly applicable, also beyond the legal domain. Although the idea of using proof standards was inspired by legal practice and legal examples are used in the article, the model of argumentation is intended to be domain independent.[11]

Let us focus here on the main original contribution of Farley and Freeman's paper, regarding its model of the burden of proof. Among the contemporary computational models of argument, only [100], [163] and [209] explicitly model burden of proof. Burden of proof, however, has several aspects:

1. the *distribution* of the burden of proof among the parties;
2. the *proof standard*, i.e. the level of the burden, which must be met to discharge the burden; and
3. the *kind* of burden (e.g. burden of production, burden of persuasion).

While all the contemporary models cited here handle the distribution of the burden of proof, Freeman and Farely's model was the first and only to model proof standards, such as preponderance of the evidence. Modeling the distinctions between the different kinds of proof burdens would have to wait for another ten years, until [213].

The degree to which the models handle the procedural aspects of argumentation vary. Farley and Freeman's model, like the model of [209], is more relational than procedural. These models may be viewed as nonmonotonic logics, with proof theories in the form of a dialogue game, analogous to the way Lorenzen [164] uses a dialogue game as a proof theory for intuitionistic logic. Although burden of proof does play a role in these models, they do not fully take into account epistemological and other resource limitations one must face when searching for and interpreting documents and evidence to construct arguments. In

[11] Personal communication via email with Kathleen Freeman.

fully dialogical models of argument, such as those of [34], [100], and [163], the *knowledge base* of rules and facts is constructed during the dialogue, along with the arguments which use these rules and facts. In relational models of argument, the knowledge base is fixed and assumed as input to the model.

Farley and Freeman initiated a line of research on modeling proof standards, taken up shortly thereafter with my own work with Nikos Karacapilidis, on the Zeno system [103] (see 7.4). Whereas Farley and Freeman's system was based on Toulmin's model of argument, Zeno was based on Kunz and Rittel's Issue-Based Information Systems (IBIS) model of argument and focused on supporting deliberation dialogues. In addition to the legal proof standards modelled by Farley and Freeman, Zeno included some proof standards specifically relating to deliberation dialogues, including "no better alternative" and "best choice".

Ronald Leenes [156] presented a critical analysis of the then existing computational models of burden of proof, in procedural models of legal argumentation, in the light of the Dutch law of civil procedure.

Henry Prakken [202] investigated the role of burden of proof in legal argument and considered to what extent existing models of burden of proof are able to handle shifts of the burden of proof from one party to another. [218] showed how shifts in the burden of proof can depend on the argumentation scheme applied to construct the argument. [219] investigated how burden of proof, and shifts of burden of proof, can themselves be the subject of dispute in a legal proceeding.

In [213] Prakken and Giovanni Sartor investigated relationships between legal presumptions and burden of proof and presented the first computational model of argument to distinguish different kinds of proof burdens, including the burden of production, the burden of persuasion and the tactical burden of proof.

In [104] I, together with Doug Walton and Henry Prakken, introduced the Carneades computational model of structured argument and burden of proof, including a model of various legal proof standards inspired by Farley and Freeman's work. The two systems have much in common, but one significant difference is that Carneades is based on Walton's theory of argumentation [276], while Farley and Freeman's model is based on Toulmin [265]. Carneades includes a model of the critical questions in Walton's theory of argumentation schemes, and shows how different kinds of critical questions have a different impact on the distribution of the burden of proof.

Prakken and Sartor continued the work of [219] in [214], where they developed a computational model of arguments about how to allocate the burden of persuasion.

Katie Atkinson and Trevor Bench-Capon [23] investigated the possibility of modeling some proof standards using different semantics for Dung argumentation frameworks [86]. For example, the scintilla of evidence standard is modeled as membership in some preferred extension (credulous acceptability) and beyond reasonable doubt was modeled as membership in all preferred extensions (skeptical acceptability).

In [105] Doug Walton and I presented an overview and survey of work up to 2009 on computational models of proof burdens and standards. That paper also presents an extension of the Carneades model of argument which distinguishes the open, argumentation and closing phases of argumentation dialogs and shows relationship between theses phases and the different kinds of proof burdens, including the burdens of claiming and questioning.

Finally, and most recently, Prakken and Sartor [217], inspired by Carneades, have shown how to extend the ASPIC+ model of argument [208] to provide similar support for proof burdens and standards.

It is a bit ironic that Farley and Freeman, coming from outside of the field of AI and Law and interested in argumentation in general, not limited to the the legal field, would be the first to recognize the importance of proof standards for dialectical procedures under conditions of incomplete, uncertain and inconsistent "knowledge", given the prominent role of proof standards in legal procedures and practice. This article, and their subsequent longer AI and Law Journal article on the same topic [93], with Freeman listed as the first author, were their only contributions to the field of AI and Law. The research was done as part of Freeman's PhD thesis at the University of Oregon [92]. Farley was Freeman's thesis advisor. Freeman's now 18-year-old daughter was born just one month after her dissertation defense. She worked as an adjunct professor at the University of Oregon for a number of years before taking time off to raise her family. She has in the meantime returned to the University of Oregon, as an adjunct instructor and, as of this year, as a career instructor and Director of Undergraduate Studies in the Department of Computer and Information Science. Farley, now professor emeritus at the University of Oregon, focused his subsequent research on the subjects of artificial evolution and algorithmic graph theory.

6.5 Edwina L. Rissland and M. Timur Friedman. Detecting Change in Legal Concepts [233]. *Commentary by Kevin Ashley*

When teaching a seminar on Artificial Intelligence and Law, one searches for research papers that address a significant topic for the theory or practice of law in a way that law students can understand but that may also impress the computer scientists in the class. This paper satisfies those constraints. It addresses the topic of concept change or conceptual drift, important for both legal theory and practice, it illustrates conceptual drift with an extended legal example - changes in the meaning of "good faith" of the debtor's proposed payment plans in personal bankruptcy law - and it employs an elegant method using a structural-instability metric to measure structural change in concepts represented as decision trees induced from cases.

Law students need to learn that legal concepts change over time. In my AI and Law Seminar, this paper reinforces a lesson they first encounter in Edward Levi's *An Introduction to Legal Reasoning* [158], excerpts of which are still assigned in some first year courses (and in my seminar). By framing concept drift in computational terms and subjecting it to empirical observation, the paper makes a significant contribution to knowledge representation in AI and Law and addresses a phenomenon relevant to machine learning in general.

In fact, this is one of the very few research papers in AI and Law that does address concept change over time. Given the theoretical and practical importance of this feature of law, it is rather surprising how little AI and Law research addresses it or makes computational use of the chronological order of legal decisions. And yet, case decisions unfold over time. Sometimes the changes are abrupt due to a landmark decision or the passage of reforming legislation. Or the changes may be gradual as judges respond to evolving social conditions and technological assumptions. Ramifications of legal decisions percolate within and across jurisdictions. Sometimes they coalesce into trends that affect such decisions as attorneys' assessments of the likelihood of outcomes and value of settlements. That is, attorneys would take trends into account if they perceived them. All of the decisions are recorded in databases: if only there were ways to monitor trends and their ramifications for what attorneys think they know about the law of a subject at any given time.

This research took an important step in that direction. The researchers created a chronologically ordered "example-stream" of 55 cases (from the BankXX project [231] and see 5.3) addressing the "good faith" issue over a span of more than ten years. Each case was represented as a vector of the case decision (i.e., whether that the debtor had [not] submitted the plan in good faith) and 61 features of the bankruptcy scenarios (e.g., the debtors, debts, creditors, employment, cash flow, proposed payments, amount of debtor's surplus, percentage repayment of unsecured debt, inaccuracies in debtor's representations, etc.). Each vector was input in chronological order to a machine learning algorithm, C4.5, resulting in decision-tree representations of the concept of "good faith" up to that point in time. In each decision tree, a node is either a leaf node indicating that "good faith" was satisfied (or not) or a test on an attribute, for instance, tests for duration of the proposed plan, amount of payments and surplus, or motivation and sincerity of the debtor. As each new decision-tree representation of the "good faith" concept is generated, it is compared to the previous one to determine if any structural changes have occurred, that is, if any attributes have moved, appeared, or disappeared. This is accomplished by comparing the values that occur at each location in the old and new trees. Concepts can be generalized by adding a disjunct or deleting a conjunct. They can be narrowed, by removing a disjunct or adding a conjunct. Or an attribute's relevance can change or its value be inverted. A structural-instability metric is then applied: All of the changes at each level are summed. A weighted sum of the changes over all levels is computed, in which changes in attributes at a higher level of the tree (closer to the root) are given more weight. This reflects the use of an information theoretic metric in which more predictive attributes occur higher up in the tree. An average of these summations of change across each successive pair of trees is then computed and examined for (statistically significant) changes in its slope or trend. A negative slope indicates conceptual instability is decreasing; a positive slope indicates increasing instability and triggers the hypothesis that the concept is drifting.

Interestingly, to test that hypothesis, the old tree is continually updated, a possible replacement tree is formed, and their predictions compared over the next 12 examples. If the new tree is more predictive, it confirms the concept's drift.

In the annals of AI and Law research, it is both fascinating and convincing when a computational metric tracks a phenomenon apparent in the "real" legal world. Ejan Mackaay's early work comes to mind on predicting outcomes of Canadian capital gains cases and identifying borderline cases that proved controversial in the tax law expert commentaries. This article is another example. The authors' explanation relating features recorded by the structural instability metric to actual developments in the bankruptcy case and statutory law of "good faith" is fascinating and convincing.

In the history of ICAIL conferences, this paper may be the "last word" on concept drift, but it should not be. Legal concepts change in other ways beside structurally. Courts or the legislature may retain an attribute but change how it is to be tested. The authors do not provide an actual example of a decision tree representation of "good faith", but the tests appear to be binary: either an attribute like the "duration of the proposed plan", or the "amount of payments and surplus", is included in the concept's definition or not. Sometimes, however, the tests are operationalized in more quantitative terms. A range of acceptable values may be specified. For instance, the workout plan should be completed within 18 months or its end points could be changed to make the test harder (12 months) or easier (24 months) to satisfy. In addition, as in Levi's extended example of the demise of the requirement of privity of contract in product liability cases, the conceptual change may be driven by the need to achieve a more coherent rule given: (1) the accumulating positive and negative examples against a backdrop of (2) trade-offs in the competing and changing policies underlying the concept and legal regulations in which it is embedded. (See Rissland's and Collin's 1986 "The Law as Learning System" [226]). A nice challenge would be to represent these mechanisms and use them, along with the structural techniques, to model conceptual change. One could tackle a history of real cases, a difficult task. Or, for purposes of intelligent tutoring, one could represent a pedagogically-inspired sequence of cases, complete with a capacity to generate arguments for and against alternative ways to accommodate changes perceived as teleologically desirable. This would make Levi's example computational. (See Rissland's 2009 "Black Swans, Gray Cygnets and Other Rare Birds" [238]).

6.6 L. Thorne McCarty. An Implementation of Eisner v. Macomber [179]. *Commentary by Kevin Ashley*

This paper, written more than a dozen years after Thorne McCarty's development of the theory of prototypes and deformations in Taxman II [174], is the most compact and readable account of that landmark work in AI and Law and its application to its most famous example, the seminal income tax case named in the title. The account of the competing example-based arguments of the majority and dissent in the U.S. Supreme Court opinion is a compelling lesson for law students eager to understand legal argument as well as for researchers eager to model it. McCarty's argument scheme accounts for aspects of real legal argument: the use of prototypical, precedential and hypothetical examples and the normative and cognitive task of drawing mappings across examples, pointing out invariant features which, one argues, justify treating them as the same. His theory of prototypes and deformations is about representing and reasoning with legal concepts; it deals explicitly with their open textured and dynamic nature. Like Edward Levi [158], he believes that the meanings of concepts in legal classification rules change as the rules are applied. As McCarty says (elsewhere), "Legal concepts are not static structures but dynamic ones: they are typically constructed and reconstructed as they are applied to a particular set of facts." Prototypes are examples that fit the definition of the concept. Deformations transform a prototype into other examples that arguably fit the definition of the concept; they preserve certain invariant properties of the concept (at the expense of others).

In the Eisner argument, each Justice represented the concept of taxable income using different prototypes and mappings. Mrs. Macomber, owner of 2200 shares of Standard Oil common stock, received a 50% stock dividend, resulting in 3300 shares. The Internal Revenue Service (IRS) imposed a tax on the distribution, but she objected that the stock dividend was not taxable income under the 16th Amendment. There were three relevant prior examples: the Lynch and Peabody cases, where the distributions of a corporation's cash or of another corporation's stock, respectively, was found taxable, and the Appreciation Hypothetical, in which, everyone agreed, mere appreciation in the value of a corporation's stock without a transfer of shares was not taxable.

According to the form of the argument, in arguing that a distribution is taxable income, the advocates (and justices) draw analogies to distributions in past cases or examples. The analogies are mappings; some aspect of the distribution in a past case is mapped onto the distribution in the current case and justifies treating the current case like the past one. Accordingly, the taxpayer must define taxable income to exclude the Eisner facts and the Appreciation Hypothetical but include Lynch and Peabody. The IRS must define taxable income to include Eisner, Lynch and Peabody but exclude the Appreciation Hypothetical. Crucially, each advocate's definition and treatment of the examples must in turn be justified by a coherent theory. As McCarty memorably stated,"The task for a lawyer or a judge in a 'hard case' is to construct a theory of the disputed rules that produces the desired legal result, and then to persuade the relevant audience that this theory is preferable to any theories offered by an opponent."

The competing arguments and McCarty's explication of them, both in terms of corporate income tax law and of the theory of prototypes and deformations, have provided much food for thought by law students and AI and Law researchers ever since. In outline, J. Pitney argued that the stock dividend was not taxable, because after the transaction, Macomber owned the same ratio of the corporation that she started with (3300 / 750,000 = 2200 / 500,000) the same as if there were no transfer at all. In modeling terms, his argument draws a mapping from the Appreciation Hypothetical prototype to the Eisner facts, preserving as invariant the shareholder's proportionate corporate ownership. By contrast, in arguing that the distribution was taxable, J. Brandeis' strategy in dissent was to show some kind of continuum linking distributions of common stock to distributions of cash so that it would be irrational to treat one as taxable and the other not. His argument draws a mapping from Lynch's taxable distribution of cash to distribution of a corporation's bonds or preferred stock to distribution of common stock as in Eisner. It preserves as invariant the property that each confers on the recipient some tradeoff between expected return of corporate earnings and risk, differing only in how much return and at what risk. If one is taxable, so should all be taxable.

McCarty's system was a serious attempt to perform this sophisticated and creative kind of legal argumentation. Designed to reason with rights and obligations, it would select prototypes, search for and construct mappings that relate the pro-side examples while excluding counter examples, and provide criteria for assessing the arguments. His Language for Legal Discourse (LLD) [177] represents concepts like ownership, distribution, shares, bonds, common and preferred stock and the rights and obligations of holders thereof. It supports defining legal concepts in terms of logical templates and as prototypes (positive and negative examples) as well as techniques for matching invariant components to the problem. Some invariants are provided (like the ConstantStockRatio that compares shareholder ownership ratios). The program had strategies and a procedure to search for other invariants with the goal of identifying some property the problem shares with the examples but not with the counter examples. It would construct the complex invariants from simpler components. Presumably, it would use the rights and obligations associated with different types of corporate distributions and some general information about the goals of corporations and shareholders to construct a complex invariant like the Equity/Debt continuum.

That similarity, of course, should serve as a legal basis for reasonably arguing that both distributions should be taxed (or not). Why does it matter legally that the pre- and post-share of corporate ownership is the same or that the distributions are all part of the continuum of risk/return? If each side could make arguments by pointing out invariants linking the case to favorable precedents, what is the basis for choosing one side's argument over the other?

This extended example has always struck me as a pedagogically effective way to focus law students in my AI and Law seminar on important questions like these. Students can grapple with the plastic nature of legal concepts, a legally realistic argument scheme for reasoning with legal concepts in terms of underlying statutory aims and principles, and the questions of how one evaluates those arguments.

The example also served well to introduce my students to the career long efforts of many AI and Law researchers to implement a computational model underlying the argument scheme so beautifully framed by this example. Since ICAIL 95, researchers are making steady progress on dealing computationally with the complexity of this and similar argument schemes, representing why certain factual features make a legal difference in terms of the underlying aims and principles of the substantive law.

6.7 Edwina L. Rissland and Jody J. Daniels. A hybrid CBR-IR approach to legal information retrieval [232]. *Commentary by Adam Z. Wyner*

Retrieving relevant legal opinions from the large corpora has long been a key problem in common law jurisdictions, where the opinions of judges are distributed amongst the courts. Legal professionals are faced with the difficult and significant task of identifying the cases from the corpora which are relevant to a problem case. Law libraries have always been central to law firms, containing extensive indexed collections of cases or collocations according to precedential relationships, e.g. shepardized cases. In the current period, large legal informatics companies, LexisNexis and Thomson Reuters, sell services to the legal industry to support legal professionals in finding relevant opinions. The need and value of successful textual search tools has always been important to the legal industry.

[232] is set in the context of research from the late 1980s to mid 1990s, where the problems and limitations of information retrieval (IR) approaches using Boolean keyword search in corpora of legal information were starting to become clearer and more pressing, and where the tools and techniques of Artificial Intelligence applied to legal information were becoming better developed, e.g. case-based reasoning (CBR). Boolean keyword search was widely used for IR, yet had a range of problems - it depended on the user knowing what terms to use, how to formulate and refine the query, and there was no assurance that relevant documents were returned (recall) or that all those returned were relevant (precision). In general, such an approach had no knowledge representation. Another approach enriched the texts in the corpus with some conceptual information [114], where cases were associated with frame information that was used for search. CBR focussed on the analysis of factual aspects of cases [234, 230]. IR and CBR were at opposite ends of the knowledge representation spectrum: IR can apply to any case in a large case base, but is shallow (only textual), does not support reasoning, and does not have a strong sense of the relevance of the documents that are returned; CBR can only apply to the cases that have been annotated, but supports reasoning, and indicates high relevance. [232] represents an effort to bring these strands together by using the knowledge representation of CBR to refine IR queries. The objective is to combine the best of both - the refined queries derived from CBR can be used to quickly search large corpora using standard IR techniques.

[232] propose and develop a technique in which a problem case is represented as a generic case frame filled in with the specific facts of the case. It outputs a set of documents considered relevant to the problem case. It does this by comparing the problem case to cases in a claim lattice, which are cases sorted according to similarity and difference of case factors along the lines of HYPO [227]. The most on-point cases are selected, from among which a 'best exemplar' is selected. The full-text of this best exemplar is fed into a processor which selects the top unigrams or bigrams and generates queries, which are then used to query a larger corpus of cases. A relevance feedback method is used to improve results: a user judges the relevance of the documents returned given initial queries; by tagging documents as relevant, this causes the query processor's weights to be altered, which returns a more refined query.

For corpora, [232] use cases that they have previously analysed, e.g. the 25 cases of [230]. These are used to derive the query. An "answer key" is created, being the cases that are being searched for; for one domain, there are 128 cases bearing on the home office deduction as identified by a keyword search. The keyword search also sets a baseline return average precision (the average of precision values from amongst the different levels of recall).

In [232] it is reported that the approach improves search results significantly over the baseline, meaning that by using the CBR case base to support refinement of the query can improve results. Using this approach, a legal professional must still evaluate the relevance of the documents returned, but it has relieved her of formulating queries and gives her access to large document collections in a problem-specific manner.

Case-based reasoning and textual information extraction and retrieval have continued to be important topics in AI and Law [279, 283]. The fundamental motivations have, if anything, increased, for there is now greater access to a wider spectrum of legal information on the Internet. Case facts, one of the basic ingredients of common law opinions, are essential to identify, organise, and compare relevant cases. However, while a variety of approaches have been applied, the core issues remain, for facts are represented in a variety of linguistic forms and represent complex knowledge. However, there has been a significant change from the 1990s that bodes well for future progress in the area - the movement to open source data (legal corpora available unencumbered and on-line) and software development (where tools are made openly available for research development). In this way, researchers are able to reuse, develop, evaluate, collaborate, and integrate research as never before. Thereby, the research community can decompose the large, knowledge intensive, complex problems of legal informatics into smaller problems and engineered solutions.

7 Melbourne 1997

In 1997 ICAIL moved for the first (and so far the only) time to the Southern Hemisphere, and was hosted by the University of Melbourne from June 30th to July 3rd. John Zeleznikow was General Chair, with Dan Hunter as Co-Chair and Karl Branting as Programme Chair. The Programme Committee expanded to fourteen: five US, seven Europeans (three from UK, two from the Netherlands, one from Germany and one from Norway) an Australian and one from Japan. Ross Quinlan was among the invited speakers. The conference also featured the first (and so far only) conference breakfast, held in Melbourne's Royal Botanical Gardens, at which Edwina Rissland gave an invited address. Six papers have been selected. While traditional concerns such as conceptual retrieval were still explored (7.5), new, technology driven, concerns were becoming important. These included ontologies (7.1), but significantly two of the chosen papers relate to the World Wide Web (7.3 and 7.4). These papers represent early adoption of the new technology: Navigator was still the standard browser and Google was not to be founded for another year. The immediate, lasting, impact of the web on AI and Law (and most other computer science topics) should not be forgotten or underestimated[12].

7.1 Trevor Bench-Capon and Pepijn Visser. Ontologies in legal information systems; the need for explicit specifications of domain conceptualizations. [37] *Commentary by Enrico Francesconi*

The importance of an ontological approach in the legal domain, introduced by the work of Valente and Breuker (see scetion 6.1 above), was later stressed by Bench-Capon and Visser [37] with regard to information management in general and for the legal domain in particular. The authors identified specific motivations for using ontologies. In particular, following Gruber's approach [111], they acknowledged ontologies as essential for sharing knowledge

[12] It also impacted strongly on how conferences were organised. At this time submission was by multiple hard copies which had to be distributed to reviewers by "snail mail". Electronic submission was still for the future: originally, controversially, as an option.

between systems, as well as for knowledge base assessment, knowledge acquisition and reuse, and, finally, for domain-theory development which is of particular interest for the legal domain.

Moreover Bench-Capon and Visser's work showed the limits of a rule-based approach for a legal knowledge system (due to the possible presence of contradictory or unspecified rules) and the need of an ontological commitment at different levels of abstraction, with explicit specifications of the conceptualization. Such commitment is identified in both Valente and Breuker's work and in the 'Frame Based Ontology' of Van Kralingen and Visser, which distinguishes between a generic ontology of all law and a statute-specific ontology containing the concepts relevant to a particular domain. Such distinction is able to enhance the reusability of knowledge.

This distinction were widely addressed in antecedent literature. For example [73], [76] and [130] pointed out that usually knowledge representation is affected by the nature of the problem and by the applied inference strategy. This key-point was also addressed by [73] as the *interaction problem*, with discussion regarding whether knowledge about the domain and knowledge about reasoning on the domain should be represented independently. In this respect [79] pointed out that the separation of both types of knowledge is a desirable feature, since it paves the way to knowledge sharing and reuse.

Hereinafter such concepts would be specifically underlined for the legal domain, as for example [71] which criticised a common tendency to indiscriminately mix domain knowledge and knowledge on the process for which it is used, addressing it as *epistemological promiscuity*.

The main value of the Bench-Capon and Visser's work was to provide an organic view on these issues for the legal domain, to identify the limits of rule-based systems without ontological commitments, as well as to indicate directions of research in ontology development, based on separation of types of knowledge, promoting knowledge sharing and reuse; directions which, at the state-of-the-art, are still valid. Subsequent years would see a great deal of interest in ontologies in AI and Law, and the story will be continued in section 10.1.

7.2 Layman E. Allen and Charles S. Saxon. Achieving Fluency in Modernized and Formalized Hohfeld: Puzzles and Games for the LEGAL RELATIONS Language [10]. *Commentary by Ronald P. Loui*

ICAIL 1997 [10] was the third place where some of us saw Layman Allen and Charles Saxon show their legal relations language based on the work of Wesley Hohfeld. The first two places were the Workshop on Computational Dialectics in 1995, and ICAIL 1995 [9]. I prefer the original ICAIL title: "Better Language, Better Thought, Better Communication: The A-Hohfeld Language for Legal Analysis." But any one of the three delivers the essential ontological punch. I would be happy to see Layman Allen describe his Hohfeld-based games every time there is a hardened deontic logician in the audience.

Deontic logic has occupied a privileged place in the logic of law for half a century. Names like von Wright, Alchourron, Hilpinen and Chisholm have lent the subject an esteemed position among philosophers. Even AI names like Donald Nute and J.J.C. Meyer have published major works on deontic logic. To many outside of the ICAIL community, "logic and law" axiomatically refers to deontic logic.

There are no doubt excellent puzzles that can be posed about norms and morals within a logic that has a way of representing obligation and its interplay with states of affairs. Mathematically nifty comparisons can be made with alethic modalities (possibility and necessity)

– modal logics are ripe for a common semantic analysis using Kripke accessibility function on directed graphs of possible worlds. None of this ever helped me represent a legal principle in the real world.

Perhaps Hohfeld's logical relations are not the complete answer for representing legal knowledge, but they open the door to the larger ontological world of relevant modalities. Hohfeld expands obligation and permission into four opposing pairs:

– Right vs. No-Right
– Privilege vs. Duty
– Power vs. Disability
– Immunity vs. Liability.

And he adds a second kind of duality, the correlative:

– Right cf. Duty
– Privilege cf. No-Right
– Power cf. Liability
– Immunity cf. Disability.

Perhaps the language could be reduced to a smaller independent subset (in the same way that one could talk only of obligation and negated obligation, and never talk of permission). But the analysis of common-sense terms with their syntactic symmetries was as satisfying to the algebraist as anything one could derive from Kripke semantics.

More importantly, the enlarged language permitted a more nuanced approach to the representation of legal attitudes among propositions and persons. One could say A had a duty to B, not just that A had a duty. And there were four kinds of entitlements and burdens, not just one.

A serious legal ontologist would probably go further, perhaps to add qualifiers and indexicals, such as State-Power, Civil-Liability, Human-Right, and so forth. But if I were adding a moral and legal reasoning ability to a common-sense AI artifact (e.g., Cyc), Hohfeld would not be a bad place to start.

Allen and Saxon went further, not by adding ontological distinctions, but by embedding the Hohfeld logic in a dialectical argument game. Both the use of a game and the dependence on dialectic were fairly novel at the time. Tom Gordon and I had both dabbled with dialectical games by 1995, but the addition of Layman Allen's voice made a stronger chorus.

Of course, Layman Allen was not just any purveyor of games as models of human phenomena. His WFF-N-PROOF and EQUATIONS games have been popular mathematical excursion board games for half a century. My chairman was not so sure about computer scientists modeling legal argument, developing things called ontologies, or being inspired by dialectic. But his son was a champion at the EQUATIONS game, so there must have been something to the endeavor once Professor Allen joined in it.

One might call the emphasis on games the result of good entrepreneurship on Layman Allen's part. But I am convinced that the best way to have broad and lasting impact with the new models of defeasible argument, and case-based reasoning from precedent, is to embed the ideas in a popular game. My choice was to base the argument game on cards, where the 8 of spades was a proposition, and the 8 of hearts was its negation. Layman Allen understood that it was an even more pleasant experience to hold rounded wooden cube-shaped dice in one's hand. And he colored them blue, red, black, and green, with gold lettering. It sure beats the logic paper with p's and q's.

The actual games that Allen and Saxon produced in their ICAIL papers were perhaps too complicated to understand fully in that format. But the commitment to a dialectical process was obvious, and the Hohfeld expansion of the deontic language was a breath of fresh air. Perhaps it will take time to appreciate the Allen-Saxon A-Hohfeld LRL games precisely because people must first appreciate the dialectical argument game, then they must adopt the Hohfeld modalities, and finally, they can join the two into a single whole. Such is the large cost of a large leap.

7.3 Ronald P. Loui, Jeff Norman, Joe Altepeter, Dan Pinkard, Dan Craven, Jessica Linsday, Mark A. Foltz. Progress on Room 5: a testbed for public interactive semi-formal legal argumentation [168]. *Commentary by Bart Verheij*

Remember 1997; we used Netscape Navigator for browsing and AltaVista for search. It was the heyday of the first browser war. Netscape released its 4.0 version in June (72% market share), Microsoft Internet Explorer following in October (12%). The internet was still an innocent business — or was it? —, with Microsoft employees planting their e-logo in Netscape's lawn, and Netscape placing their dinosaur on top.[13]

It was against this historical background — Wikipedia and Google didn't exist yet — that Loui presented his Room 5 system at ICAIL 1997 in Melbourne. Room 5 was developed by Loui and a team of students in collaboration with Norman, a Chicago-based lawyer, and was designed as an interactive web-based system, in which users could argue legal cases. The goal was in particular to facilitate discussion of pending Supreme Court cases, 'precisely because of the interest that members of a broad community might have in arguing them' [168]. Although the term Web 2.0 wasn't yet invented, Room 5 had an underlying collaborative community vision similar to that movement,[14] while also going a step further: the community using the tool should help find out which kinds of logical constructs and conceptual distinctions were useful and actually used, cf. the project's ambitions [168]:

1. To identify a community of web-users willing to play semi-formal legal argument games;
2. To gauge the willingness of such users to be subject to the constraints of various formats, gauge their general understanding of constructions permitted, and determine the practical limits of a few formats' expressiveness;
3. To permit a community of non-naive contributors to construct an ontology for U.S. federal law and a database of semi-structured arguments.

Each of these ambitions is still a worthy aim.

Room 5 also proposed an innovative graphical argument format based on nested tables, instead of the more common tree structures. For instance, consider an argument concerning the issue whether John is punishable with one pro-reason, namely that John has stolen a CD, and one con-reason, that he is a minor first offender. Figure 1 (left) shows the argument in a classical tree format, here in the Reason!Able program [97], while Figure 1 (right) shows the same argument in Room 5's table-based format.[15] The example shows that, in Room 5, a box encapsulated inside another represents a supporting reason, and a box next to another represents an attacking counter-reason.

By its use of nested boxes, Room 5 does not readily allow for the graphical representation of what Pollock famously refers to as an undercutting argument, i.e., an argument that attacks the connection between a reason and its conclusion. In fact, the example argument

[13] http://en.wikipedia.org/wiki/Browser_wars#The_first_browser_war.
[14] http://en.wikipedia.org/wiki/Web_2.0.
[15] These and other graphical formats for the presentation of arguments are discussed in [269].

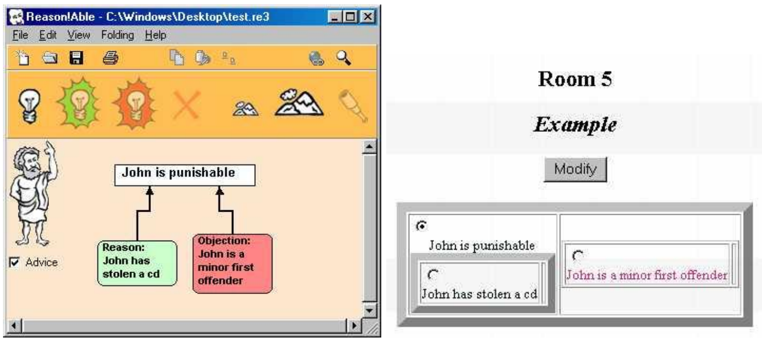

Fig. 1 An example argument in Reason!Able and in Room 5

Fig. 2 An undercutting argument in ArguMed

may be better represented as such an undercutting argument, since the fact that John is a minor first offender only blocks his stealing as a reason justifying punishment, and does not imply that John is *not* punishable, since there can be another, independent, reason for John's punishability. Instead of using a graphical representation, Room 5 uses a text-based conditional format for the representation of undercutting arguments (r THOUGH p THUS NOT(q)), citing the paper, 'where the argument is used only to attack an argument for q rather than to establish NOT(q)'. One style of graphical representation of undercutters is by the use of nested arrows, e.g, as in ArguMed [269] (Figure 2).

The study of argumentation tools has progressed significantly since 1997, both in terms of systems investigated, cf. the overviews [152], [185], and [248], and in terms of research infrastructure, cf. the journal *Argument & Computation* and the biennial international COMMA conference series on the computational modeling of argument. Argumentation support tools also have proven their commercial value, with Reason!Able's successor Rationale[16] a telling example.

7.4 Thomas F. Gordon and Nikos Karacapilidis. The Zeno Argumentation Framework [103] *Commentary by Katie Atkinson*

The Zeno argumentation framework [103] was proposed as one of the first attempts to define a formal model of argumentation that could specifically be used to structure discussions in

[16] http://rationale.austhink.com/.

online consultation systems for e-government. The field of AI and Law had already contributed to research on computational models of argument, but Zeno was proposed at a time when interest in argumentation for e-government applications was in its infancy.

The power of the World Wide Web to act as an enabler of new forms of democratic participation by the citizenry was only just beginning to be recognised in 1997 when the Zeno paper was published. Zeno was developed as part of research on a European project, GeoMed, whose motivation was towards supporting the process of negotiation and mediation during geographical planning. As such, the main idea behind Zeno was to use the Web to enable interested citizens and representatives of public interest groups to access, view and discuss geographical plan proposals set out by governmental administrators more easily. The key innovation of Zeno was to make use of formal argumentation to structure the information relevant to the debate of concern, and so enable it to be reasoned about more rigourously by showing, for example, the dependencies between arguments and the problem solutions that the acceptable arguments entail.

The Zeno framework takes its chief inspiration from Rittel and Webber's Issue-Based Information System (IBIS) [239]. However, informal models such as IBIS were augmented in Zeno to provide a syntax for 'dialectical graphs' and a 'semantic labelling' function to support a kind of inference. The dialectical graph is concerned with capturing procedural argumentation by focussing on the role and function of the speech acts used in argumentative exchanges between individuals. The example debate used in the paper to demonstrate the framework is one in which a husband and wife discuss what kind of new car they should purchase.

The dialectical graph representing such a discussion contains two kinds of node, *positions* and *issues*. If a position is disputed, it gives rise to an issue, and then reappears in the graph as a choice with respect to that issue. Positions may either be *factual statements* or *preference expressions*, in which one position is preferred to another. Edges are of one of four types: the children of issues are either *choices* (factual positions) or *constraints* (preference expressions), while the children of positions are either *pros*, supporting the position, or *cons*, attacking the position. An *argument* is a pair of positions, one of which is a pro or con of the other. Now, given such a graph, it is possible to calculate the status of an issue, with respect to a range of proof standards common in the AI and Law literature (e.g. [88] see section 6.4). Thus, the kind of inference supported by Zeno means that the debate participants can determine, at any given point, which are the currently winning and losing positions in the debate (which may well change if further arguments can be found and introduced by the participants).

The Zeno framework that was set out in [103] was taken forward into a number of different projects in which the framework was further developed. For example, Zeno has been used in e-democracy pilot applications in a number of subsequent European projects, such as the Delphi Mediation Online System (DEMOS) [169], which was developed to offer innovative Web-based services facilitating democratic discussions and participative public opinion formation.

Zeno made a significant step in bringing to the attention of the AI and Law community the potential for improving government policy consultations and discussions through the use of formal, structured models of argument that can underpin such tools. Nonetheless, the challenge still remains of seeing such tools being brought into widespread use; the most ubiquitous online tools used by governments to gather public opinion as part of policy making can probably be characterised by the e-petitions Web sites[17], which permit only the use of unstructured information that means that sense-making and analysis remain cumbersome activities. However, researchers are addressing the many open challenges of delivering supportive yet easy-to-use systems for e-democracy. It is well recognised that the burden of learning how to use a structured tool should be minimised as far as possible to ensure that the citizenry is willing and able to make use of such applications with as little training as possible [171]. It is in this spirit that systems such as Parmenides [74] and Compendium[18], to name a couple of representative examples, have been developed.

The Parmenides tool was developed as a way to enable citizens to critique government policy proposals that were presented in a structured form of argumentation, yet understanding of this structure was not required; users simply had to provide yes/no answers to a series of questions posed about a policy proposal and its justification. The Parmenides tool is currently undergoing significant extensions as part of the European IMPACT project[19], in which an innovative argumentation toolbox for supporting open, inclusive and transparent deliberations about public policy is being developed. The toolbox will contain four separate but interacting tools, including a policy modelling tool being developed by Thomas Gordon and colleagues, based on Gordon et al's Carneades system [104]. The Carneades system is software developed for constructing, evaluating and visualising arguments from formal models of legal concepts, rules, and cases. Its development as part of the IMPACT project shows how the goals of the Zeno framework continue to be pursued through the development of state-of-the-art tools for computational argumentation for use in e-government.

Since working on the Zeno system, Nikos Karacapilidis has gone on to investigate computer-supported argumentation and decision-making in a number of large scale projects, such as the HERMES system [146], which similar to Zeno, is based on the theoretical foundations of computational argumentation. More recently, Karacapilidis has worked on the CoPe_it! system [147], which is a Web-based tool designed to support collaborative argumentation and decision making, again following the legacy of the Zeno framework.

Although, as highlighted above, many challenges remain in realising the goal of deploying e-government systems that structure information to make it more useful to both citizens and policy makers, the Zeno paper brought to prominence the recognition that formal argumentation can have a useful and significant role to play in this endeavour.

7.5 J.C. Smith. The Use of Lexicons in Information Retrieval in Legal Databases.[259]. *Commentary by Erich Schweighofer*

The contribution of Joe Smith for the ICAIL 1997 in Melbourne, Australia, was his last in a long series of excellent papers of the University of British Columbia Faculty of Law Artificial Intelligence Research (FLAIR) Project. For over a decade, FLEXICON (and later FLEXLAW) was a constant presence in AI and law research.

The FLAIR group did some research in logic and law but became much more famous for the achievements in the field of "Best Match - Non-Boolean search engines". Like all experts on legal information retrieval, Smith shared the view - supported by research such as [60] - that Exact Match Boolean is not a suitable choice for the user and more efficient support is greatly needed. A high-level play with Boolean word combinations based on an

[17] See the UK site: `http://epetitions.direct.gov.uk/`, or the US site: `https://wwws.whitehouse.gov/petitions`. Accessed April 3rd 2012
[18] See: `http://compendium.open.ac.uk/institute/`. Accessed April 3rd 2012
[19] See: `http://www.policy-impact.eu/`. Accessed April 3rd 2012.

excellent knowledge of about 1 million words in several million documents is something for the expert searcher, not the lawyer as such. In 1997, search engines were predominately Boolean with a ranking algorithm. Important improvements were in development (like the later very successful PageRank of Google) but were not available then. However, the basic concept of a reduced meta representation of a textual document for improving retrieval and ranking in this meta environment had already existed for a long time. Representations can be complex (like Carole Hafner's LIRS (Legal Information Retrieval System, see 2.2) or workable (e.g. the norm based thesaurus of Jon Bing, see 2.4). The best developed system was and still is the FLEXICON information system of Joe Smith.

The FLEXICON information system consists of three databases: the linear text of the original documents, multiple lexicons, and a lexical representation of each document, called a FlexNote, a (mostly) machine created abstract of each document. The lexicons consist of legal concepts, cited cases, cited statutes, proper names and facts. The FlexNote consists of case header information and four quadrants containing the most significant concepts, facts, case citations and statute citations in the case. The search screen has a very similar structure. A lookup feature allows online consultation of all lexicons.

The automatic computation of FlexNotes depends very much on the quality of the lexicons and the programmes for extracting relevant terms and phrases. The Legal Concept Dictionary as a domain lexicon of words or phrases used by legal professionals had to be constructed by hand. It is based on stems. Synonyms or semantically similar concepts are linked. Cited cases and statute citations are recognized through template matching. Fact phrases are extracted using term distribution and proximity information with a lexicon of "noise words". Such words typically point to relevant facts before or after their occurrence. Joe did not report all details of this method but it much have taken a tremendous effort to fine-tune the rules in order to achieve reported results. As a strong supporter of the"best match" concept, Joe used the Vector Space Model of Salton *et al* [242] for relevance ranking. Both documents and queries are represented as vectors in a multidimensional space. The dimensions are terms extracted from the document. Relevant documents were retrieved by comparing queries to document profiles based on the Cosine formula. Document terms are weighted with inverse document frequency. An additional feature is user-based weighting of the importance of a term, using categories of High, Medium, Low or Not. Empirical tests of FLEXICON using a small database showed higher recall and precision than Boolean retrieval. Further, very relevant cases got a higher ranking. It is sad to state, however, that - to my knowledge - the intended FLEXLAW commercial application was never launched.

In contrast to many other projects, Joe Smith got the chance to develop FLEXICON to a nearly commercial product. His work on linking a text corpus to a semantic representation is still, after so many years, the strongest in coverage even if methods may be more advanced now. For my own research on semiautomatic text analysis, I got a major impetus from his work. The main ideas of Joe - start with a strong thesaurus, refine with (semi)automatic text analysis, and rank and classify document using a reduced representation - are still present. Obviously, other methods were developed, such as that of Jody Daniels (with Edwina Rissland) - case-based reasoning combined with the INQUERY information retrieval engine ([232] and see 6.7); and Marie-Francine Moens - automated content retrieval and extraction ([180] and see 11.2). In recent years, others authors have started similar research projects. The link of a meta representation to a text corpus is still not solved sufficiently. None of advanced information retrieval, machine-learning and neural networks has be able to close efficiently the knowledge acquisition gap for conceptual representations (the same is true also for logic-based representations). The basic concept of a meta representation with (semi)automatic extraction of this knowledge is still the best approach today.

Jon Bing and Carole Hafner were the pioneers in working on conceptual (now called ontological) representations of legal text corpora. Joe Smith refined the methods, added the component of (semi)automatic knowledge acquisition to this research and moved a very long way towards a nearly commercial product. The response of legal authors is also here decisive. They have to consider using (semi)automatic text analysis tools for grasping and analysing the legal content. Then, sufficient knowledge and energy would be available for a highly supported use of conceptual information retrieval and (semi)automatic text analysis tools. However, legal language is considered so much as a core of legal competence and has achieved such high quality and efficiency that some replacement of this work by software will still take many years. It is somehow a vicious circle: without sufficient resources no quality and with no quality no competitive advantage for this method. Thus, more convincing work has to be done in order to add this method to the tool set of lawyers.

7.6 Vincent Aleven and Kevin D. Ashley. Evaluating a Learning Environment for Case-Based Argumentation Skills [6]. *Commentary by Trevor Bench-Capon*

One of the major achievements of AI and Law has been the exploration of case-based reasoning which began with Edwina Rissland [225], was continued by her in HYPO (with Kevin Ashley) and CABARET (with David Skalak), was further progressed by Kevin Ashley as CATO (with Vincent Aleven) and IBP (with Steffie Brunighaus), and which Kevin continues to this day in work such as [109]. CATO, which is the topic of [6] is probably the most influential of these, in the sense that it is taken as the model of case-based reasoning in work such as [212], [41], [78] and, most recently, [135] and [285].

Although many have taken it as representative of an approach to case-based reasoning, CATO was in fact directed towards a quite particular task. Carried out in the Learning Research and Development of the University of Pittsburgh, CATO was, as the title of [6] clearly indicates, intended to support *learning*, specifically the learning of case-based argumentation skills. The particular skill CATO was intended to teach law students was *how to distinguish cases*. Not every difference between cases can be used to distinguish them - it is important that the difference makes the current case more favourable for your side than the precedent was. Moreover there are counterarguments against potential distinctions: it may be that some other feature can be used to take away the significance of the difference (to *downplay* it, as [6] puts it).

This task explains the major differences between CATO and Ashley's own PhD program, HYPO [16]. Whereas HYPO used *dimensions*, features of a case which could range from an extreme pro-plaintiff value to an extreme pro-defendant value, CATO used *factors*, which are, effectively, particular points on a dimension which always favour either the plaintiff or the defendant. For discussions of the differences between dimensions and factors see [236] and [39]. The point of this simplification in CATO was to allow the students to concentrate solely on the cases as bundles of factors, and so not be diverted by considerations as to whether dimensions applied, or which side they favoured. The other major innovation in CATO's representation of cases was the use of a *factor hierarchy*. This related the base level factors which corresponded to points on HYPO's dimensions, as children of more abstract factors. The presence of a child would provide a reason for or against the presence of the abstract parent factor. This hierarchy could then be used to argue that a different factor could substitute for, or cancel, a factor offered as a distinction if it related to the same abstract factor, thus providing a mechanism for downplaying. Finally a feature of [6] is the emphasis it placed on a set of eight basic argument moves, intended to provide heuristics for students looking for good distinctions.

These changes from HYPO enabled CATO to provide a teachable method for finding good distinctions in a body of precedent cases, and the evaluation among law students (in passing it should be noted that the thoroughness of the evaluation of CATO is praiseworthy, and typical of AI and Law done at Pittsburgh under Kevin Ashley, although unfortunately less typical of AI and Law work in general) suggested that it was effective. CATO's widespread influence, however, comes from the fact that these three features are also very attractive to those who wish to analyse case-based reasoning using logical tools. The representation of cases is terms of factors, means that we already have the cases expressed as intermediate predicates (see [160] and the discussion in 4.1). Thus the difficulties in reasoning from world knowledge to intermediate predicates noted in 4.1 are obviated: systems based on factors can focus on the more tractable issues associated with reasoning from intermediate predicates to legal consequences. This facilitated the development of logics for case-based reasoning (as exemplified by CATO) in, for example, [212] and [135].

Cases described in terms of intermediate factors were also the necessary starting point for the theory construction of [41] and its empirical exploration in [78]. In addition that work exploited the factor hierarchy since that hierarchy provided a good way of relating factors through associated values, by mapping abstract factors to values. This is the mechanism by which precedents can be applied to cases comprising different sets of factors in [41].

Finally the explicit set of argument moves was very useful in the design of dialogue systems such as those in [78] and in devising argumentation schemes for case-based reasoning as in e.g. [285]. Thus the features that made CATO such a useful support for learning case-based reasoning techniques, also made it highly suitable as a source of inspiration for those modelling case-based reasoning, specially those attempting to use logical tools. The question arises, however, as to whether there are other case-based argumentation skills that should be investigated, as well as those relating to distinguishing cases.

CATO was intended to support the teaching of some particular case-based argumentation skills, and it certainly achieved this. But it is not only the classes at Pittsburgh in the mid-nineties which have benefited from it, but in a very real sense a large proportion of the AI community, myself included, have learnt much of what they know about reasoning with legal cases from studying CATO[20].

8 Oslo 1999

The Seventh ICAIL was held at the University of Oslo from June 14th to 18th 1999. Oslo at this time of the year has very long days, not quite the midnight sun, but twilight seamlessly becomes dawn. Jon Bing was Conference Chair and Tom Gordon was Programme Chair. The Programme Committee continued to expand, this time reaching eighteen members

[20] This paper also introduced one of the classic AI and Law cases, *Mason v Jack Daniels Distillery*, 518 So.2d 130, 1987 Ala. Civ. App., to AI and Law. A bartender, Tony Mason, invented a cocktail, Lynchburgh Lemonade comprising Jack Daniel's whiskey, Triple Sec, sweet and sour mix, and 7-Up. It proved surprisingly popular. Mason met Winston Randle, a sales representative for Jack Daniel Distillery, and they talked about the drink, and its possible use in a promotion. Approximately one year later the defendants were developing a national promotion campaign for Lynchburg Lemonade. Mason claimed that he had parted with the recipe because he had been told that his band would be used in the promotion. In fact Mason received nothing. The jury found for the plaintiff, but awarded only a dollar in damages. Following [6] Mason joined *Eisner v Macomber* (see 6.6) and the wild animals cases following *Pierson v Post* (see 5.1) in the AI and Law canon.

(eight from the US, seven from Europe, two from Australia and one from Japan). Among the invited speakers was Peter Johnson who, since his presentation of [141] at ICAIL 1991, had, with his company SoftLaw, made a commercial success of legal expert systems in Australia, and there was a panel discussing why there were not more examples of successful commercial exploitation. Only one paper (8.1) has been selected from this conference, but by bringing together abstract argumentation frameworks and dialogue semantics, it represents an important bridge between the argument and dialogue work of the nineties, and that which was to come in the first decade of the twenty-first century.

8.1 Hadassa Jakobovits and Dirk Vermeir. Dialectic semantics for argumentation frameworks [140]. *Commentary by Trevor Bench-Capon*

This paper by Hadassa Jakobovits and Dirk Vermeir [140] represented the only contribution to ICAIL of these two authors. Jacobovitz, whose PhD thesis formed the basis of the paper, published (as far as I can discover) nothing subsequently: her supervisor Dirk Vermeir has published over a hundred items, mostly in logics for AI venues. The paper is very untypical of ICAIL, being very much in the the logics for AI genre, with a very formal presentation and three theorems. None the less it concerns topics central to AI and Law, which had begun to be discussed over the previous decade and were to be increasingly a focus of interest in the next, namely argumentation and dialogue. Also it continually motivates its formalism with reasonably realistic examples drawn from law, establishing the relevance of the formalism, and the motivation for providing this high degree of rigor. It presents the idea of formal abstract argumentation framework as introduced by Dung [86], and uses this formalism to define the notion of a dialogue type. This in turn allows the definition of strategies. The authors then use this apparatus to examine two proof theories, where a proof theory is determined by a dialogue type and a winning criterion. One theory is based on *useful argument* dialogues and applied to an example where we have a conflict and a dispute as to which of two legal principles should take precedence. The other is *rational extension* dialogues, which they illustrate with an example from family law.

The importance of this paper is that it clearly shows how an elegant formal presentation can illuminate and clarify issues that all to easily got lost or conflated in informal presentations. It does, however, raise the question: why is there so little work like this in ICAIL? The discussion of Hall and Zeleznikow [127] in 9.4 will draw further attention to the general lack of evaluation in ICAIL papers. There is some empirical work, and some formalisation, but very few results in the form of theorems. This is in stark contrast to other AI conferences such as IJCAI, AAAI, ECAI and AAMAS where the majority of papers positively bristle with theorems. And it is not that the AI and Law Community does not contain people who do this sort of work: several of its leading lights go to AI conferences and present highly formal work complete with theorems. I think it is something to do with the interdisciplinary nature of ICAIL: as noted in the introduction, sessions are always plenary and so one writes an ICAIL paper for a mixed audience of lawyers and computing people, knowing too that the latter group will contain both applications people and experimenters as well as theoreticians. The result here is that stress tends to be placed on the motivation rather than the results: that the formalisation is useful for considering a genuine problem in law is essential at ICAIL: it is considered undesirable to simplify the problem to facilitate the proofs. Contrast perhaps some of the work on normative reasoning in agents systems such as [2]. Moreover, for such an audience, *that* something can be shown is at least as interesting as *how* it can be shown, whereas the technical detail of the mathematics may be what is interesting for a specialist audience.

9 St Louis 2001

The Eighth ICAIL took place from May 21st to 25th 2001 at Washington University, St Louis with Ron Loui as Conference Chair and Henry Prakken as Programme Chair. The Programme Committee increased to twenty-one members (eleven US, eight Europeans and two Australians) and there was an attempt to diversify the membership by bringing in some expertise from outside the traditional AI and Law community. This was quite effective in widening the scope of the conference and two of the selected papers (9.1 and 9.2) originate from a commercial rather than an academic environment. Of the other two, one (9.3) reflects long standing concerns of argumentation, use of precedents and legal coherence, whereas the other (9.4) addresses the practical need for more evaluation of AI and Law work. Ron Loui was a superb host, and the St Louis conference was (for me at least) the friendliest conference, and the divisions between case-based and rule-based approaches were finally reconciled. Highlights included a visit to the historic Courtroom where the Dred Scott[21] case had been initially tried in 1847, and the conference dinner. The dinner was held in a rooftop lounge and the pre-dinner drinks were held on the terrace as the evening sun went down with views over the city including the Gateway Arch. The Banquet speaker was listed in the programme as Senator Barack Obama, but unfortunately he had to withdraw at the last moment. Best of all was the spontaneous dancing to the *impromptu* music of Marc Lauritson on piano and Tom van Engers on pocket cornet.

9.1 Jack G. Conrad, and Daniel P. Dabney. A cognitive approach to judicial opinion structure: applying domain expertise to component analysis [80]. *Commentary by Paul Thompson*

Twenty-five years ago, when the inaugural International Conference on AI and Law was held, the field of artificial intelligence (AI) was receiving much attention both in the media and in research settings in academia and industry. Early approaches to AI, beginning in the 1950s, which sought to achieve artificial intelligence through general high level logical principles, had largely given way to the new paradigm of expert systems. While expert systems were developed in many disciplines, including law, the field which received the most attention was medicine. Expert system researchers held that AI could be achieved by modeling the heuristics and domain knowledge of experts in a narrow field, such as diagnosing blood infections. Many AI and Law researchers, on the other hand, were still influenced by earlier approaches to AI based on more general principles of logical and legal reasoning. The dominant approaches in 1987 were Case-based Reasoning (e.g. [227]) and Expert Systems e.g. [262]). That these approaches remained a constant theme of AI and Law can been seen from the papers discussed in this article, and these approaches were still dominant in 2001.

Working at West Group[22], Conrad and Dabney had access to legal editors and editorial processes which were not available to the same degree to other AI and Law researchers. While other researchers continued to develop the case-based reasoning and expert system

[21] Dred Scott was a slave who sued for his freedom. The case eventually reached the Supreme Court, where Scott lost 2-7.

[22] Now part of the Thomson Reuters Corporation.

approaches, Conrad and Dabney were able to study the legal editorial environment in a more comprehensive way. One of their long term goals was to develop automated techniques that could enhance the human editorial process. Conrad and Dabney succinctly described their study as follows.":

> Empirical research on basic components of American judicial opinions has only scratched the surface. Lack of a coordinated pool of legal experts or adequate computational resources are but two reasons responsible for this deficiency. We have undertaken a study to uncover fundamental components of judicial opinions found in American case law. The study was aided by a team of twelve expert attorney-editors with a combined total of 135 years of legal editing experience. The scientific hypothesis underlying the experiment was that after years of working closely with thousands of judicial opinions, expert attorneys would develop a refined and internalized schema of the content and structure of legal cases. In this study participants were permitted to describe both concept-related and format-related components. The resultant components, representing a combination of these two broad categories, are reported on in this paper. Additional experiments are currently under way which further validated and refine this set of components and apply them to new search paradigms.

Conrad and Dabney noted in their paper that case law does not have logical overall structure at the corpus level, nor at the individual case level. The hypothesis of [80] was that experienced attorney editors had learned what the components of case law documents were and that the editors could reach a consensus on these components. If this were true, the authors believed that it would be possible to build AI tools, e.g., expert systems, which would use this consensus knowledge representation of case law components to aid in the case law editorial process. I was a colleague of Conrad and Dabney at West at the time these experiments were done. Although I was also involved with AI and Law experiments, I worked in a different part of the company and so was not then familiar with their work. In recent years, however, I have begun new research on computational analysis of biomedical journal articles. In this field there is currently active research on understanding and using the component parts of scientific articles [272] and [220]. Conrad and Dabney's paper provides a model for this research.

Although I am not quite as close to the field today as then, it seems clear, as described in [139], that West has built on the research reported in [80] in the development of its next generation of legal search. Thus [80] remains an important paper because the old issues and debates of AI have not been resolved. Search and computational linguistics have been increasingly seen as fields for the application of statistical machine learning techniques with no grounding in semantics or discourse. There is a focus on big data. For domains, such as law and medicine, where human domain expertise is still an important part of the editorial environment, but where the volume of information is overwhelming manual approaches, it is important to develop mixed automatic and human approaches that can provide high quality editorial enhancement in a cost-effective manner.

9.2 Khalid Al-Kofahi, Alex Tyrrell, Arun Vachher and Peter Jackson. A Machine Learning Approach to Prior Case Retrieval [4]. *Commentary by Alex Tyrrell*

Sadly, on August 3, 2011, the Artificial Intelligence and Law community lost a friend and visionary leader, Peter Jackson. Peter was Chief Scientist and head of the Corporate Research and Development department at Thomson Reuters. Throughout his tenure at Thomson Reuters, Peter was known for his innovative thinking, leadership, mentorship and most of all, his endearing friendship. For those who worked closely with Peter, his accomplishments and legacy are still seen every day in our continued efforts to provide the best technology solutions to meet the needs of legal practitioners.

Looking back over Peter's body of work, including his books, scholarly publications, and patents, a series of papers from 1998-2003 stand out. During this time, Peter led a research effort aimed at assisting legal editors as they perform the task of tracking and identifying history relationships between case law documents. In this context, there are two types of history: direct - decisions impacting prior cases in the appellate chain - and indirect or commentary affecting cases outside of the appellate chain - e.g., declined to follow another opinion. The system, called History Assistant, used numerous NLP, information retrieval and machine learning techniques to link court opinions and identify direct history, within the appellate chain, as well as identifying cases that provide indirect history.

The first paper related to History Assistant appearing in ICAIL was entitled "Anaphora resolution in the extraction of treatment history language from court opinions by partial parsing," [3]. This paper described the groundwork for much of the information extraction and semantic processing phase of History Assistant. In 2001, the follow-up paper, "A Machine Learning Approach to Prior Case Retrieval" [4], described a more developed system for retrieving prior cases with direct history, by combining multiple innovative technical solutions. Finally, at the culmination of the research on History Assistant in 2002-3, Peter was first author of a journal article in Artificial Intelligence [138], which described the final system in full detail.

As mentioned, [4] focused on retrieving prior cases from within the appellate chain of an instant case. This proved to be exceedingly difficult for many reasons. Foremost, court opinions may have multiple prior cases, and the goal should be to retrieve all of them. Similarly, cases may be argued, in whole or in part, in different courts, perhaps multiple times. This creates complex linkages in the appellate chain. Finally, key information like party names may change, or are otherwise highly ambiguous, e.g., State v. Smith.

To solve these challenges a number of techniques were used. These included extracting history language from the opinions using a semantic parser, estimating prior and joint probability distributions between courts in the US (e.g., Supreme Court decisions have a high-likelihood of having a prior case), performing extraction and resolution of party names, combining structured and unstructured search, and classifying candidates using an SVM. At the time, the combination of NLP, IR and machine learning methods in this domain was not only novel, but given the requirements of near-perfect recall (97.9-99.2% reported) and precision at or above 60%, the performance bar was set very high as well. Importantly, as an applied research problem, the goals also included satisfying the workflow requirements of our editorial staff by providing a robust, high performance system that was convincingly beneficial, compared to the traditional approach.

To be sure, History Assistant was an ambitious and innovative example of applied research. Ultimately, rather than become a single, dedicated platform, History Assistant proved to be more useful as a technology incubator, leading to further refinement and deployment of certain underlying technologies, to meet other needs.

On a personal level, I began my professional career working on History Assistant, beginning in the late 1990s. This work led to one of my first publications, namely [4]. Now, over a decade later, after taking leave twice for graduate school and a postdoctoral fellowship, I continue to work and do research in the Research and Development department at Thomson Reuters. In addition, many of the original contributors to the History Assistant project have remained to face the challenges confronted by the corporate Research and Development group. Under Peter's guidance and leadership, we have collectively brought many innovative products into the legal marketplace, including the ResultsPlus document recommendation engine on Westlaw and most notably WestlawNext, launched in 2010.

Reflecting back, one has to admire Peter's dedication to History Assistant. He often remarked that it was one of his last research efforts as the primary technical contributor. Peter was quite proud of the accomplishments we achieved, and never wavered in his belief that our research and development must take on difficult problems.

9.3 Jaap Hage. Formalising Legal Coherence [120]. *Commentary by Michał Araszkiewicz*

In [120], Jaap Hage presents an account of a coherence theory of legal justification and theory construction and focuses on the relation between abstract goals and concrete regulations. In consequence, this contribution offers insights into the methodology of theory construction in AI and Law research, teleological reasoning and the application of coherence theory to the field of legal justification. In 2001 all these topics were intensively discussed in the field of AI and Law and this stemmed from, *inter alia*, seminal papers by Donald Berman and Carole Hafner (see [46] and 5.1), which emphasized the role of values in CBR and Thorne McCarty (see [179] and 6.6), which argued for seeing legal reasoning as theory construction and theory application. For instance, several papers presented at the 2000 Jurix conferences dealt with this subject [38], [201], and [119]. Hage had himself addressed the problem of teleological reasoning at the 1995 ICAIL [117]. Hage's analysis of the logical mechanism of the HYPO system [16] also constitutes a part of the context of this paper [120]. General jurisprudence is, however, another important part of the context (in particular, Robert Alexy's theory of legal principles as optimization commands, cf. [7]). The collaboration of Jaap Hage with Alexander Peczenik (resulting in joint papers [125], [193] and [126]) also might have stimulated Jaap's focus on the role of the concept of coherence in legal reasoning. Peczenik was one of the most dedicated advocates of application of coherence theories to the law. See [192], which has a preface by Hage, for the most authoritative statement of Peczenik's theory.

The paper advances a formalized theory of an important aspect of legal reasoning, namely, the comparison and choice of different regulations on teleological grounds. The theory may be seen as an extension of Reason Based Logic, developed by Hage in his earlier work (beginning with [116]). In contrast to the earlier work, however, the theory defended in this paper allows for comparison of alternative conclusions as well as balancing reasons for and against a particular conclusion. In the theory several specific predicates (for instance, concerning the relation of contribution of a regulation to a goal or detraction of regulation from realizing a goal) are introduced and defined. However, as well as the formal representation of an important part of legal argumentation the general philosophical and legal theoretical perspective is an important contribution. Hage begins with the general epistemological discussion of foundationalism and coherentism (invoking also Rawls' theory of reflective equilibrium). Then, he introduces an early version of his own account of coherence which he refers to as *integrated coherentism* (it must be noted that integrated coherentism is a theory built at a general philosophical level and not exclusively at a legal-philosophical level). Further, the paper contains a critical analysis of Dworkin's theory of legal interpretation which is rightly assessed as too vague for computational purposes. Then Alexy's theory of legal principles is briefly presented as a point of departure for Hage's theory of weighing teleological reasons that plead for and against different regulations.

This paper, as one of the first AI and Law contributions I had encountered, played an important role in shaping both my interest in legal epistemology and in the methods for dealing with the problems of legal reasoning. The paper offers a persuasive perspective on the connection between the general philosophical, jurisprudential and formal (computational) issues in modeling of legal reasoning. Also, it was one of the contributions which fostered my own research on legal coherence and my interest in balancing values in legal reasoning (e.g [14]).

The ideas engaged with in the paper were further developed by Hage, as regards both the philosophical underpinnings of the theory and the formal framework. Hage's idea of integrated coherentism was refined in e.g. [121]. The formal theory of comparing alternatives in the law was also further elaborated in [121]. Moreover [120], has itself been cited by such authors writing on teleology in AI and Law such as Trevor Bench-Capon, Henry Prakken and Giovanni Sartor.

Hage's paper [120] is an example of a contribution in which the ideas are introduced and partially described in detail, and it is very interesting to follow the author's later work to see how these ideas have been developed. However, this paper is not only a kind of first step in the line of research - it contributed seriously to the discussion concerning modeling of teleological reasoning around the beginning of this century. The paper offers a set of formal schemes representing different configurations of reasons pleading for and against a given regulation and therefore it can be used as a source of inspiration (or a direct basis) by researchers who are interested in formalizing teleological reasoning in the law, in the continental law culture in particular. [120] is also a relatively rare example concerning the connections between general philosophical and jurisprudential research on the one hand and AI and Law formal methods on the other.

9.4 Jean Hall and John Zeleznikow. Acknowledging insufficiency in the evaluation of legal knowledge-based systems: Strategies towards a broad based evaluation model [127].
Commentary by Jack G. Conrad

To the best of our knowledge, [127] represents the first self-reflexive work from and about the ICAIL conference. The point the authors make is that there is a deficiency of reported evaluation in the field of legal knowledge-based systems. Yet the field of legal knowledge-based systems is not alone in this. There's been a similar deficiency in general knowledge-based systems as well.

The operative word in [127] is *evaluation*. In varying degrees, this work covers three related areas corresponding to three separate levels of granularity: "conventional" software systems, knowledge-based systems (KBS), and legal knowledge-based systems (LKBS).

The paper collectively makes at least three meaningful contributions to the field. First, it provides a thorough review of the literature and sketches a picture of the existence (or lack thereof) of evaluation in software, KBS and LKBS papers. Second, it performs an analysis of the presence of evaluation in works published by ICAIL and compares the three then most recent conferences (1995-97-99) with those of the first ICAIL (1987) to see how the presence and use of evaluation had changed (not much). And third, the paper presents a number of existing evaluation-related models, such as

- fundamental models ([61]; [1]; ISO/IEC 14598 and 9126, *On Describing Software Evaluation Methods*);
- traditional verification and validation-related models ([144]; IEEE Standard 729-1983, *On Defining Software Verification and Validation*);

- the O'Keefe and O'Leary hierarchy of expert system quality [187]; and
- the Capability Maturity Model from CMU [190].

In addition, Hall and Zeleznikow challenge the community to improve its practices in the area of evaluation and to maintain a heightened role for it going forward.

Coming from the more established and arguably more mature research communities like Information Retrieval (SIGIR) and Knowledge Management (CIKM), I was surprised by the number of published ICAIL papers that have little if any energy devoted to evaluation topics. Far too many fail to answer this most basic question: *okay, here is your model, design, system - does it work, and if so, how well?* The Hall and Zeleznikow paper addresses this topic head-on, in a detailed, methodical manner. In other communities, submitted works that perform no evaluation or fail to discuss an evaluation process would be disqualified. We need to take the message of their work to heart. As a result, as both authors and reviewers, we should strive to ensure that evaluation is an essential component of our research reports and weigh strongly the presence or absence of an evaluation component in other submitted works.

Hall and Zeleznikow have held the light up to the AI and Law community and its context in the broader KBS field and shown its warts, pimples, blemishes - its evaluation-related deficiencies - for all too see. If it was not clear before this work, it should be now. If author-researchers wish to convince their audiences of the utility of their approaches and systems, they need to be prepared to present suitable performance metrics illustrating that they do what the authors claim they do. In this sense, from this point forward the threshold for acceptability has been raised.

Although appreciation for the role of evaluation in KBS may have grown over the past decade, and as a result, the presence of evaluation in KBS and LKBS papers may have increased commensurately, the message of the Hall and Zeleznikow work concerning the need for critical evaluation of models, designs and systems is as important today as ever before.

10 Edinburgh 2003

The Ninth ICAIL took place at the University of Edinburgh from June 24th to 28th 2003. It was held in the Law School rather than Computer Science, and so took place in the historic Old College (1789, designed by Robert Adam). Conference Chair was John Zeleznikow (who had moved from Melbourne since 1997) and Co-Chair was Lillian Edwards. Giovanni Sartor was Programme Chair. The Programme Committee now reached twenty-nine members, (twelve US, fifteen Europe and two Australia). Katia Sycara was among the invited speakers. The continuing and growing importance of the WWW is reflected in the associate conference workshops which included workshops devoted to the law of electronic agents, e-government, on-line dispute resolution and web-based information management. Two of the papers selected also reflect this trend: one (10.1) concerns ontologies, by now very closely associated with the web, and the other (10.2) is inspired by research on multi-agent systems, a technology greatly facilitated by the web. The third is a paper concerning case-based reasoning, another in the line going back to HYPO, a thread which runs through the whole series of ICAIL conferences.

10.1 Alexander Boer, Tom M. van Engers, and Radboud Winkels. Using Ontologies for Comparing and Harmonizing Legislation [62] *Commentary by Enrico Francesconi*

In later years ontologies, introduced to AI in Law in work such as [267] and [37] (see sections 6.1 and 7.1) have been widely used for modelling, comparing and harmonizing legal knowledge. In this context the work of Boer, van Engers and Winkels [62], together with others in the same period, can be considered as an organic attempt to use a Semantic Web approach for describing the semantics of legal resources, based on XML to structure documents, and RDF/OWL to represent legal knowledge within a multilingual and pluri-jurisdiction scenario. The work of Boer, van Engers and Winkels describes the Metalex initiative, within the E-POWER project, aimed at defining a jurisdiction-independent XML standards for legal documents and an ontology able to provide a jurisdiction-independent vocabulary of legal concepts, conceived for comparing and harmonizing legal knowledge in different jurisdictions. The ontology here described stressed the subtlety of norms, their roles to regulate the reality of a specific domain, as well as the role of an ontology to represent a normative system, resolve conflicts between norms and provide instruments for the harmonization of norms

In this respect an essential reference for legal knowledge modelling, viewed from a more philosophical point of view, is the work of Sartor [245] for his contribution to the characterisation of legal concepts so as to provide a formal model of their structure and a logical framework able to deal with the specificity of legal reasoning.

Though at an embryonic stage, Boer, van Engers and Winkels's work showed the use of Semantic Web technologies for modelling legal knowledge and resources, thus paving the way to succeeding works on legal knowledge modelling and acquisition within the same legal Semantic Web framework.

10.2 Katie Greenwood, Trevor Bench-Capon and Peter McBurney. Towards a computational account of persuasion in law [110]. *Commentary by Henry Prakken*

By 2003 a number of models of case-based legal argument were established in AI and Law. In [110] Katie Greenwood (now known as Katie Atkinson[23]) and her PhD supervisors Trevor Bench-Capon and Peter McBurney introduced the idea that deciding a legal case is not primarily a reasoning problem but a decision-making problem. They proposed to model a judicial decision as "an *action*, not the derivation of some *fact* about the case". They thus saw legal decision making as a form of what philosophers call *practical reasoning*. In AI practical reasoning had been studied in models of planning and of rational agency, but in AI and Law no one had yet explicitly regarded legal decision making as a form of practical reasoning.

The basis of Greenwood's approach is a presumptive argument scheme (cf. [275]) inspired by the well-known practical syllogism *Agent P wishes to realise goal G, if P performs action A, G will be realised, therefore, P should perform A* and the argument scheme from good consequences. *Action A of agent P brings about G, G is desirable, therefore, P should perform A*. Greenwood's scheme introduced a distinction between goals (which can be fully realised or prevented) and societal values (which can only be promoted or demoted to certain degrees). This distinction has now been widely accepted as important.

[23] Atkinson is Katie's married name. She was on honeymoon in the Maldives during ICAIL 2003.

In the Current Circumstances R
we should perform action A
to achieve New Circumstances S
which will realize some Goal G
which will promote some value V

As usual in an argument-scheme approach, Greenwood defined various ways to attack application of this scheme, categorised as denial of premises, alternative action for the same effect, negative side effects of the action, and interference for other actions for other goals or values.

Greenwood then applied her general model to CATO-style case-based reasoning [5]. Briefly, the current circumstances R are subsets of CATO's base factors, the possible actions A are deciding for plaintiff or for defendant, the new circumstances S are that the case has been decided for plaintiff or for defendant, goals G are conjunctions of factors and one of the possible outcomes in S and values V are attached to the abstract factors of CATO's factor hierarchy. For example, the abstract factor Questionable Means was related to the value Dishonesty is Punished. Greenwood then discussed how the resulting model accounts for CATO's eight argument moves.

In some respects Greenwood's proposal was still preliminary. For example, the use of the argument scheme was not yet embedded in formal models of argumentation-based inference. Also, the application to CATO-style case-based reasoning was perhaps not the best possible illustration of the general model. For example, the knowledge representation involved considerable (re-)interpretation of the cases and factor hierarchy in terms of values. Also, the modelling of goals as conjunctions of factors and outcomes is somewhat unnatural, since as observed by Greenwood, actions are thus guaranteed to achieve the goal. Furthermore, it might be argued that the explicit modelling of both goals and values as required by Greenwood's argument scheme is somewhat restrictive, since in many practical arguments only one of these is made explicit.

Nevertheless, it is undeniable that many arguments presented in court and also many legal policy arguments (such as on proposals for legislation) involve goals and values. In my opinion, therefore, the presentation of this paper at ICAIL 2003 was an important event in AI and Law, since it opened up a new line of research in which formal and computational models of practical reasoning can be applied to important forms of legal reasoning. In addition, the paper and its successors (such as [21]) were also influential in general AI, since they gave one of the first AI accounts of practical reasoning as a form of argumentation [22].

One still open problem is a fully satisfactory formalisation of argument schemes for practical reasoning in formal models of argumentation-based inference. One issue here is accrual of reasons or arguments, since actions will typically realise or prevent several goals and promote or demote several values. One approach is to use a general formalisation of accrual of arguments (as done in [42] which uses [205]). Another option is to reify statements about realising/preventing goals and promoting/demoting values in a first-order language so as to allow for arguments about sets of goals and values (as done in [44]).

10.3 Stefanie Brüninghaus and Kevin D. Ashley. Predicting Outcomes of Legal Cased-Based Arguments [72]. *Commentary by Trevor Bench-Capon*

This paper [72] introduced IBP (Issue-Based Prediction), the latest in a distinguished family of programs. IBP is a child of CATO, a grandchild of HYPO, a niece to CABARET and a cousin of BankXX, all of which have been mentioned often earlier in this paper. What is distinctive about IBP is that whereas its ancestors generate arguments and leave it to the user to decide which they will accept, IBP chooses between the two sides, and so predicts the outcome of the case.

Although primarily a development of CATO (see 7.6), and using the US Trade Secrets cases analysed by Vincent Aleven (now Steffie's husband), IBP borrowed an important feature from CABARET (see 3.4). IBP uses a "logical model" - a logical and-or tree based on the Restatement of Torts - to structure its reasoning into issues, just as CABARET uses statutory rules to provide its top level structure. The leaves of the and-or tree split the consideration of a Trade Secrets Case into five issues. Then, just as CABARET uses case-based reasoning when the rules are exhausted, IBP uses case-based techniques to resolve these issues. The values for the leaves are then be propagated up the tree to produce a prediction. In some case where factors for both sides were present it was found necessary to ascribe strengths to different factors in order to resolve the conflicts appropriately. Therefore IBP identifies *knock-out factors*, whose presence is almost always enough to be decisive for their side and *weak* factors, where the importance depends to a great extent on the context provided by other factors. Sometimes, when there are conflicts that cannot be discounted, IBP will abstain, but the empirical work showed such cases to be very rare. Note how the consideration of a case becomes increasingly less holistic as we descend from HYPO. In HYPO factors (or rather dimensions points) were homogenous; in CATO they were grouped through the factor hierarchy so that some factors could cancel out and substitute for others. In IBP they are partitioned into issues for independent resolution.

An admirably thorough evaluation of IBP was reported in [72], and it proved to give a correct prediction in over 90% of cases, outperforming a number of other standard machine learning algorithms. This is a gratifyingly high, but rather similar to other techniques developed specifically in an AI and law context, such as neural nets ([28] and see 5.6: disappointingly no neural net comparison was used in [72]), the argumentation based theory construction of Alison Chorley [78] and the argument mining approach of Maya Wardeh [278]. Some reasons why cases are misclassified are given; that a case may have features which make the decision *sui generis*, that the decision is simply wrong (fortunately rare) that the set of factors do not capture all aspects of the case, or that the ascription of a factor may be a matter of debate.

An important point is remember that the starting point for IBP is cases represented as *factors*. Thus the identification of the intermediate predicates has already been done, and this identification is often where the difficulties lie (see 4.1 and 13.1). In part the high degree of success can be explained by fact that the move from facts to factors is part of the prior analysis. IBP, however, was only part of Steffie's work: she also produced SMILE, a program intended to identify factors in text. Thus the grand aim was to combine SMILE and IBP to predict the outcomes of cases from textual descriptions such as the squib - a short description intended to include a synopsis of the important facts. The combination is reported in [19]. Again there is a very thorough evaluation, which serves to illustrate the difficulty of the task set to SMILE, and the consequent deterioration in predictive performance when the starting point is the textual description rather than the pre-analysed case. SMILE+IBP gives only a 70% success rate (compared to a baseline of 66%). But none of this should detract from IBP, or from Steffie's contribution: rather we should take it as a clear demonstration of just how important the analysis of cases to identify intermediate factors is, and how very hard that problem is to automate.

11 Bologna 2005

The Tenth ICAIL went from the historic University of Ediburgh, founded in 1583, to the even more historic University of Bologna, the original *alma mater*, founded in 1088, and probably the oldest University in the world. Giovanni Sartor was Conference Chair and Anne Gardner was Programme Chair. The Programme Committee shrank a little to twenty six (twelve US, eleven Europe, two Australia and one Israel). Cristiano Castelfranchi was an invited speaker. The Bologna conference was also remarkable for the number of associated workshops. Several of these reflected the spread of potential applications of AI and Law: E-Commerce, E-Government, Dispute Resolution, the law of Electronic Agents, Trust, and Legal Education. Other workshops looked at specific aspects of AI technology: ontologies, argumentation and data mining. There were a number of interesting papers, but the one selected here (11.1) addresses text summarisation. This topic had become important, and has since become still more important, as the quantity of legal texts available had grown so rapidly with the WWW. Then - and even more so now - automatic processing seems essential to cope with the volume of accessible material.

11.1 Ben Hachey and Claire Grover. Automatic legal text summarisation: experiments with summary structuring [113]. *Commentary by Frank Schilder*

Summarization of legal text is different from summarization of news messages; directly applying techniques developed by the NLP summarization community has shown itself to lead to unsatisfactory results. The retrieval and summarization of case law documents, however, is important because lawyers need to find precedent cases quickly [184]. Most of the approaches to summarization of legal text at the time Hachey's and Grover's research in [113] was undertaken were based on rule-based extraction systems. This work was one of the first approaches to analyzing legal text with the help of machine learning techniques.

Similar to legal text, scientific text contains different argumentative 'zones' that need to be recognized in order to create an adequate summary, as noted by Teufel and Moens [264]. Hachey and Grover transferred this concept of analyzing complex text to the legal domain similar to Farzindar and Lapalme [89] who crafted rules for the segmentation of English and French Federal Court of Canada judgments.

The main contribution of [113] was to introduce machine learning approaches to legal summarization based on argumentative zones. In addition to a battery of different machine learning algorithms (SVMs, Decision trees etc.), they achieved superior results by using maximum entropy Markov models. They further improved the set of features by harnessing a linguistic annotation pipeline. The theory of argumentative structure also allowed for customized summaries with varying structure and the possibility of personalization.

I became aware of Hachey and Grover's work when I started working for Thomson Legal and Regulatory in 2004. Peter Jackson, the VP of the R&D department at the time, asked me to investigate approaches to summarization of legal documents. Based on Hachey and Grover's work we developed a prototype for summarizing court decisions and we also had the opportunity to have Ben Hachey, a native Minnesotan, present his work to the group in Saint Paul. Their annotation work had a big impact on our internal annotation work and their approach influenced our subsequent in-house summarization work. Most importantly, we were excited to have Ben Hachey start working for our R&D department in December 2011.

The AI and Law community has appreciably benefited from this work by an increased interest in machine learning approaches applied to the processing of legal text. The work encouraged researchers in this area to try methods that had been formerly used mainly on news text to venture into more complex text types such as case law documents. Some of these efforts include the following:

- Francesconi and Peruginelli [91], which describes a machine learning approach to categorization of legal documents in ten areas of practice (environmental, administrative, constitutional, etc). It cites [113] as an early application of machine learning to legal AI.
- Galgani and Hoffman [95]. This paper reports on legal citation classification in the context of summarization. It draws on Hachey and Grover's discussion of the relative complexity of legal arguments with respect to scientific publications to convey the nature of task.
- Mochales and Moens [181] and see 13.2. They follow Hachey and Grover in using statistical classifiers for argument analysis, but, they identify argument mentions instead of classifying sentences by argument type.

Hachey and Grover also provided a publicly available corpus of annotated UK House of Lords judgments that can be used for comparing new systems with their work. There is a trend in the Computational Linguistics and NLP community to base scientific results on such publicly available data collections in order to ensure reproducibility. [113] also utilized an important concept for deeper text understanding by applying argumentative zones. Future work will surely build on this concept and address additional questions such as how machine learning techniques can be applied to other legal collections of different types (e.g. statutes). Researchers will also turn their attention to methods that allow bootstrapping techniques in order to port systems to new legal domains more rapidly. Finally, more work needs to be done to create abstracts instead of extracts as summaries; in addition, the readability of the generated summaries needs to be improved by creating more sophisticated discourse smoothing techniques.

All of these exciting new avenues can draw their inspiration from the work on summarization of legal text created by Hachey and Grover in [113].

12 Stanford University 2007

The Eleventh ICAIL was held in California, at the Law School of Stanford University from June 4th to 8th 2007, with Anne Gardner as Conference Chair and Radboud Winkels as Programme Chair. The Committee now reached thirty-three members (eleven US, eighteen Europe, two Australia and one each from Mexico and Israel.) The importance of the WWW continued to be evident with workshops on legal ontologies, e-discovery, e-commerce, online dispute resolution and the semantic web for law, as well as an invited address from Deborah McGuiness and sessions on related topics such as agents and text mining. Two of the selected papers are also related to this technology, one on e-discovery (12.4) and one on opinion mining from the then recent phenomenon of law blogs (12.3). The other selected papers relate to aspects of legal argumentation: one considering the use of stories to assess evidence (12.2) and the other describing an approach to arguing about the burden of persuasion (12.1).

12.1 Henry Prakken and Giovanni Sartor. Formalising Arguments about the Burden of Persuasion. [214] *Commentary by Douglas N. Walton*

Burden of proof is clearly very important in legal argumentation because cases can be won or lost depending on which side the burden of proof lies. But there is more to it than that. In order to be able to evaluate legal argumentation, it is necessary to understand how burden of proof and the closely related notion of presumption work as essential devices of argumentation in legal cases. Recent research in artificial intelligence and law has provided the first clear and well worked out logical framework using an argumentation model. This paper ([214]) is part of a sequence of papers by Henry Prakken and Giovanni Sartor, and builds on their previous work on formal systems of argumentation, in conjunction with some other research in the field of artificial intelligence and law. It provides an argumentation-based model of burden of proof and presumption that is still being refined. The results of this research are of wide significance, going beyond the field of artificial intelligence and law, as it is important not only in evidential reasoning in law itself, but in any field where argumentation is used to analyze and evaluate evidence.

The central core of this research of Prakken and Sartor is that it distinguishes three kinds of burden of proof: the *burden of persuasion*, the *evidential burden*, usually called in law the burden of production of evidence, and the *tactical burden*. The burden of persuasion is fixed and does not change during a trial. Once it is met, it determines who wins the trial. The evidential burden operates when the case has come into the hands of the jury, although a ruling by the judge may dispose of the issue on the spot without leaving it open to the jury. The tactical burden, which shifts back and forth during the trial, is decided by the advocate by assessing the risk of losing on that issue if he presents no further evidence.

As mentioned above, the contribution of this particular paper [214] needs to be understood in light of a sequence of papers, and here we can only highlight some of the most noteworthy developments (see [216] for a survey of much of this work). Freeman and Farley [93] (see also section 6.2) presented a computational model of dialectical argumentation that included burden of proof as its key element. In particular, they defined some standards of proof that specified levels of support for satisfying a burden of proof. For example, to satisfy the preponderance of the evidence standard, the arguer has to find at least one defensible argument that outweighs the other sides rebutting arguments. The scintilla of evidence standard is satisfied by the finding of at least one defensible argument supporting the claim. To meet the dialectical validity standard, the proponent has to find at least one defensible argument, and has to defeat all the other sides rebutting arguments. Variations on these standards turned out to be vitally important in moving research on burden of proof forward.

Prakken, Reed and Walton [219] provided a formal dialogue game for persuasion dialogues in which the burden of proof can become the topic of the dispute. Prakken and Sartor argued in [213] that presumptions can be modeled as default rules in a nonmonotonic logic. In this system, it was shown that invoking a presumption can fulfill a burden of production or persuasion while it shifts a tactical version to the other party. The Carneades Argumentation System [104] showed how proof standards make it possible in a formal dialectical system to change the burden of proof during a dialogue as it progresses from stage to stage and offered a solution to an important problem for argumentation generally, namely the problem of how a burden of proof can shift from one side to the other when a critical question matching in argumentation scheme is asked by one party.

The ICAIL 2001 paper [214] presented an argument-based logic for reasoning about allocations of the burden of persuasion. It allowed for reasoning about the burden of persuasion within the logic itself, as opposed to the earlier research in which the allocation of the burden of persuasion was fixed. One example was about the ownership of archaeological goods in Italy, which illustrated well how presumption is linked to burden of proof as the dialogue moves forward to its conclusion in a legal case. In a subsequent paper [215], the previous work on presumptions and burden of proof was extended by studying the force of a presumption once counter-evidence has been offered.

This research continues to move forward, but has certainly gone beyond the point where anyone could dispute that it provides a computationally useful formal model for assisting in determinations of burden of proof and presumptions in legal argumentation. The wider implications of these findings are still not widely studied or known outside the specialized community of computer scientists in the field of artificial intelligence and law. That is why this paper, and the rest of the sequence discussed here, are so important today. Work on the subject continues, and one development that is especially important is the effort to incorporate the notions of burden of proof and presumption modeled by Prakken and Sartor into a working computational system that has a user interface that can be used in practice and sp shown to be helpful to legal professionals.

12.2 Floris Bex, Henry Prakken and Bart Verheij. Formalising Argumentative Story-based Analysis of Evidence [50]. *Commentary by Douglas N. Walton*

Prior to this paper [50], the two dominant approaches to reasoning with legal evidence in artificial intelligence and law had been argumentation and abductive reasoning (inference to the best explanation). This paper provides a formalization that combines these two approaches, by merging a formal structure of evidential reasoning with an analysis of explanation based on scripts and stories (anchored narratives). On this approach, a story can be seen as a connected sequence of events and actions of a recognizable type that fits together into a recognizable pattern. This approach fits very well with the kind of evidential situation typical in criminal cases where the prosecution has one so-called story, or account of what supposedly happened, and the defense has a different story. Evaluating the argumentation in a case by comparing the two stories seems to be the most natural way to carry out this task, and empirical evidence suggests that it is the method that juries essentially use [191]. As the authors show by means of an example of a criminal case, using their approach an adjudicator can weigh the two competing stories to see which one is the most consistent, which one is the most plausible, which one has the most gaps that suggest critical questions, and which one is based on and best supported by the evidence that is known in the case. This last factor is particularly important, as it is essential, in order to prevent errors, for an adjudicator to evaluate the story that is based on stronger evidence more highly than the competing story that is more plausible but based on weaker evidence. Clearly, however, judging this last factor requires an argumentation system that can be used to support the explanation-based story system. By showing how to do this in a formalized system, this paper has accomplished something that is turning out to be more and more powerful as a tool for artificial intelligence and law, in the research carried out since [50]. Some of this later research has been comprehensively surveyed in [47], where the results of [50] turned out to be important for the subsequent advances.

One of the nicest things about the hybrid approach is that it enables an analyst of legal evidence to provide computer-assisted visualization of an explanation that is comparable to an argument map [51]. Even better, the diagrams in [51] can combine argument and explanation by showing how each of the nodes (events or actions) or arcs (transitions from one node to another) can be supported by evidence or attacked by counter-evidence. This feature was the basic idea of the anchored narratives approach, whose advocates showed it could be applied in a helpful way to analyzing and evaluating evidence in criminal cases [273]. Another development was the results showing how stories can be investigated and evaluated in a dialogue game structure, a natural setting for explanations [53]. Such a dialogue structure uses critical questions matching an argumentation scheme to probe into the weak parts in a story, for example one put forward by a witness being cross-examined in a trial. This interesting feature of stories was further investigated in [270], extending the hybrid theory. The formal structure of the hybrid theory was further refined in [52].

Providing a formal model in [50] laid the groundwork for this subsequent research and so was a big step forward that is certainly bound to strongly influence future research, not only in artificial intelligence and law, but in many other areas as well, where the argumentation model and the story-based explanation model can be combined and applied. So far the theory appears to be not as widely known outside the artificial and law community as it should be, but that will change.

12.3 Jack G. Conrad and Frank Schilder. Opinion mining in legal blogs [81]. *Commentary by Jochen L. Leidner*

Opinion Mining and the closely related field of Sentiment Analysis were still arguably in their infancy in the domain of text analysis in 2006 when the research for [81] was conducted. Formal works exploring sentiment analysis applied to general applications (movie reviews) or specific domains (financial, consumer reviews for particular products) had only begun to be published with any frequency two or three years before.

Conrad and Schilder's paper was the very first work of empirical research to apply the new techniques of sentiment analysis and opinion mining to documents in the legal domain. In particular, this investigation tracked the opinions of legal practitioners on an assortment of legal research tools. In addition, this work was among the very first to perform an analysis of documents coming from social media and the Web, that is, from the "blogosphere" (also known as the "blawgosphere" in the legal field). To help ensure that some of their experiments were reproducible, the authors leveraged an existing sentiment analysis tool from a credible research lab, Alias-i, and its associated natural language toolkit, LingPipe[24]. Also distinct was the granularity of LingPipe's sentiment analysis analyzer. It performed character-based analysis in contrast with token-based analysis.

The paper showed that, by using standard metrics of information retrieval such as precision, recall and the F-measure, along with a human-tagged 'gold' test data set, one could measure the utility of sentiment analysis in a vertical domain like law in a practical way, for instance, tracking the opinions of law students or other customers of legal research tools.

The research first pursued in [81] in the legal field has fostered more sophisticated follow-up work combining the resources above with those of, for example, document summarization, in a ICAIL 2009 paper by Conrad, Leidner, Schilder and Kondadadi [82]. This work effectively opened up a new realm of legal text to the AI and Law community, with material quite different from the standard legal documents produced by courts for use by the community.

The work spearheaded by Conrad and Schilder was important to the field of AI and Law on several levels. In addition to being the first work of its kind applying opinion mining to texts in the legal field, it was also the first to investigate sentiment recorded in social media

[24] http://alias-i.com/lingpipe/.

such as Web blogs (or 'blawgs'). In addition, their work demonstrated that seminal, novel work has and continues to originate from within the AI and Law community. Subsequently, it has inspired more ambitious and far-reaching work and has been widely cited within the AI and Law community, but more significantly outside of the community, in areas of sentiment analysis, social media analysis, and summarization.

For a comprehensive bibliography that encompasses works cited in the Conrad and Schilder paper and provides a thorough investigation some later developments, see Pang and Lee's Sentiment Analysis compendium [189].

12.4 Jason R. Baron and Paul Thompson. The search problem posed by large heterogeneous data sets in litigation: possible future approaches to research [25]. *Commentary by Dave Lewis*

[25]In November 2005, I had my first conversation with Jason Baron and was introduced to the fascinating world of electronic discovery (e-discovery), a topic which has since come to be the focus of both my consulting and my research. For many researchers in AI and Law, however, the Baron and Thompson paper at ICAIL 2007 [25] was the first flag of warning and opportunity in advance of the tsunami that has become e-discovery.

In December 2006, the US Federal Rules of Civil Procedure were amended to supplement the term "documents" with references to electronically stored information (ESI), and to establish procedures for lawyer attention to this data. Many heads were thus extracted from many sand banks to confront the reality, described by Baron and Thompson, that millions, perhaps billions, of e-mail messages and other digital objects might need to be searched for responsive material in routine civil litigation matters.

The Baron and Thompson paper conveyed not just the need for tools to attack this problem, but what these tools would face in terms of the historical context of prior adoptions (and neglects) of information retrieval (IR) and AI in the law. In 2012, tools from IR and AI are widely used in e-discovery, particularly in the form of supervised learning algorithms targeted at improving quality and reducing cost of review.

Most importantly, [25] conveyed the central role that random sampling and statistical evaluation would play in e-discovery. In 2012, this is beyond obvious. E-discovery practitioners and vendors routinely use (and misuse) statistical evaluation. Judges discuss, and order, sampling. I recently testified in US Federal Court on recall, precision, confidence intervals, and other issues once considered arcana of information retrieval and machine learning research. But in 2007, as Baron and Thompson quite accurately describe, these issues were almost completely ignored in both the law and technology of discovery. The authors were prescient indeed to emphasize the central role evaluation would come to play. Among other things, they presented in detail the lessons of the first TREC (Text REtrieval Conference) Legal Track, introducing this effort for the first time to the ICAIL audience.

Baron and Thompson's crystal ball was not uniformly clear. Elusion, a measure they highlight, has in my opinion seen considerable misuse in e-discovery (partly because Baron and Thompson's caveats are often ignored). Overall, however, the paper gave a compelling vision of new, exciting, demanding field. In addition to the practical impacts described above, and the multibillion dollar industry around them, e-discovery has become an exciting research area within both IR and AI. A growing research community has come into being,

[25] This sub section was originally published by the author and is reprinted here by his kind permission. Thanks are due to the author for granting this permission.

presenting at forums such as the Discovery of Electronically Stored Information (DESI) workshops that have several times co-located with ICAIL, the TREC Legal Track, and the SIGIR 2011 Information Retrieval for E-Discovery (SIRE) workshop. A special issue of the journal *Artificial Intelligence and Law* in 2010 highlighted some of the most notable of this work, and we can look forward to much more to come.

13 Barcelona 2009

The Twelfth ICAIL was held at the Universitat Autonoma de Barcelona from June 8th to 12th 2009, in the splendid *Casa Convalescncia*. Conference Chair was Pompeau Casanovas and Programme Chair was Carole Hafner. The twenty-nine members of the Programme Committee comprised nine from the US, seventeen from Europe, two from Australia and one from Israel. There was a good supporting programme of workshops, on e-discovery, ontologies, privacy in social networks and modelling legal cases. Two papers have been chosen from this conference. One (13.1) revisits the concerns about how much common sense knowledge is needed to move from facts to a case described in legal terms, and the other (13.2) represents another look at the increasingly urgent issue of automatically extracting useful information from a large text corpus.

13.1 Kevin D. Ashley. Ontological requirements for analogical, teleological, and hypothetical legal reasoning [17]. *Commentary by L. Thorne McCarty*

As discussed in section 5.1 above, Donald Berman and Carole Hafner presented their paper on teleological reasoning at ICAIL in 1993 [46]. Their main point was that a robust theory of case-based legal reasoning must contain "a deeper domain model that represents the purposes behind the rules articulated in the cases." They illustrated this point in two different areas of the law: (1) the wild animal cases that are traditionally taught at the beginning of the first-year Property course in American law schools, specifically, the first three cases in Casner and Leach's casebook [75], *Pierson v. Post*, *Keeble v. Hickeringill* and *Young v. Hichens*; and (2) the worker's compensation cases that formed the subject matter for Karl Branting's system, GREBE [65]. For some reason, the worker's compensation cases have not played a prominent role in the subsequent literature, but the wild animal cases have supported a cottage industry of computational modeling, see, e.g., [30]. Then, in 2002, in a remarkable display of serendipity, a California court decided the case of *Popov v. Hayashi, 2002 WL 31833731 (Ca.Sup.Ct. 2002),* in which the plaintiff, Alex Popov, tried but failed to catch the baseball from Barry Bonds' seventy-third home run, and the judge in the case cited *Pierson v. Post* and *Young v. Hichens* in the course of his decision. Shortly thereafter, this case, too, entered the literature on AI and Law, see, e.g., [282].

In [17], Kevin Ashley takes seriously Berman and Hafner's quest for a "deeper domain model" to represent "the purposes behind the rules," and translates this quest into a search for an *ontology* that would support a Socratic dialogue in a typical law school classroom covering the cases from *Pierson v. Post* through *Popov v. Hayashi* and beyond. It is a very complex and sophisticated paper. As a running example, it includes a sample dialogue on the wild animal cases, in six parts, and four tables listing (i) Cases/Hypotheticals, (ii) Principles/Policies, (iii) Factors and (iv) Proposed Tests (i.e., Hypotheses) for the domain in question. It has an extensive list of references, which are cited copiously throughout the text, and it would serve as an excellent resource for anyone who wanted to understand the current state of research in the field. But it is primarily a “challenge” paper, as is the original paper by Berman and Hafner. In fact, by laying out the ontological requirements so clearly and in such detail, the paper raises doubts, in my mind, about whether this is still a goal worth pursuing, at least with this line of cases.

One important contribution of the paper is the identification of three roles that an ontology should support. According to Kevin, the first role, supporting case-based comparisons, “is nearly within reach of current AI and Law technology [p. 9].” This claim is supported in Section 4 of the paper, primarily by listing the relevant factors and policies in the four tables, in the style of HYPO and CATO. The real challenge arises with the second and third roles that Kevin identifies: distinguishing deep and shallow analogies, and inducing/testing hypotheses. Parts 5 and 6 of the sample Socratic dialogue were constructed to illustrate these roles, and Kevin argues (correctly) in Section 5.2 that hypothetical tests involving principles and policies are necessary ingredients in real legal arguments. But just imagine the domain ontology that would be required to handle the following instantiation of Kevin’s second role:

> It must, however, be possible to see a “baseball” as a kind of quarry like “fish” or “fox” and “putting in one’s pocket” as a kind of interception. [p. 8]

or the following instantiation of Kevin’s third role:

> One could describe all of the cases dealing with animal quarry and (even very valuable) homerun baseballs without necessarily introducing “economic goals”, and yet it is an obvious (to human reasoners) way in which to abstract a rule from these diverse cases. [p. 9]

This would not be just an ontology of factors and cases [281], but an ontology of all human (and animal!) activities and interactions.

Hence the question: Does it make sense to continue down this path? In answering questions like this in the past, I have found it helpful to distinguish between two possible goals of AI and Law research. If our goal is primarily theoretical, then we should ask whether our computational model — *qua* computational model — is likely to provide any insights into legal reasoning that we could not have achieved by a traditional jurisprudential (i.e., verbal) approach. Do we really need to build this system, or have we learned enough just by writing the paper, as Kevin has already done? On the other hand, if our goal is primarily practical, then we should ask instead whether our techniques will scale up to a realistic size. Today, this usually means that our data structures must be computable automatically (or at least semi-automatically) from the raw source material. Conversely, it does not make sense to spend years developing a domain ontology that will only tell us something that we already know, as a theoretical proposition, and can only be constructed, by hand, from the primary legal texts. Unfortunately, the wild animal cases may fall into this latter category.

Let’s face it: *The Law is AI-Complete!* This means, by analogy with the theory of *NP-Completeness* [96], that given any problem X in AI there exists a polynomial-time mapping from X into some problem Y in a particular legal domain. In order to make progress in this field, we need to identify tractable problems within AI whose solutions, when applied to the law, yield either genuine theoretical understanding or real practical applications, or both.

13.2 Raquel Mochales and Marie-Francine Moens, Automatic detection of arguments in legal texts [180]. *Commentary by Floris Bex*

Mochales Palau and Moens discuss the techniques involved in what they call argumentation mining, that is, automatically detecting the argumentative structure of a document. This highly original and novel research provides important insights which are important for a number of different fields, amongst which are argumentation and discourse theory, natural language processing (NLP) and (legal) document retrieval.

Argumentation is one of the core elements of almost any legal text. Court judgements, witness testimonies and government green papers all contain arguments for and against claims. Being able to automatically detect the structure of these arguments in legal texts would be a huge step forward for systems that handle legal document retrieval. If, for example, we want to search for a case that argues in favour of abortion (e.g. *Roe v. Wade*) we can perform a smart search by asking the system to *give me all the legal arguments in favour of abortion*. As IBM's Watson project has shown quite recently (http://www-03.ibm.com/innovation/us/watson/), such question-answering techniques and systems present an exciting development in AI, one that would be very well suited to a domain like law, where over the years vast amounts of arguments have been posed and countered.

Even though argumentation mining is an important and interesting topic, the research on this type of automatic discourse analysis is relatively sparse; this is probably due to the difficulty of NLP in general. Mochales Palau and Moens mention foremost the work on Argumentative Zoning ([263]; [113] and see section 11.1), in which documents are divided into zones with differing meanings such as 'background' and 'aims'. Other work on automatic argument analysis includes, for example, [241] which presents a system for automatically extracting argumentative structures from simple advice and warning texts (e.g. in instruction manuals). The results from this research show that argumentation mining is still very much in its infancy, particularly argumentation mining from larger, more complex texts such as legal judgements.

In order to mine arguments from natural language texts, Mochales Palau and Moens argue, a description (theory) of argumentation is needed that provides exactly and only the elements required for argumentation mining, namely a clear description of the minimal units of argumentation and the possible relations between these arguments (e.g. linked or convergent structures [94]). Interestingly, despite the computational aims and background of Mochales Palau's and Moens work, they draw solely from informal argumentation theory for their description of arguments: argumentation schemes [277] are used to type single-step arguments consisting of premises and a conclusion while the formalisation of more complex structures (e.g. chains of arguments) is drawn from pragma-dialectics [87].

Two corpora are used in [180] to test the initial formalisation of argument structures: the Araucaria corpus, which contains a structured set of arguments analysed from magazines, newspapers and so on, and the ECHR corpus, which contains a structured set of legal texts from the European Court of Human Rights. Determining arguments and their elements was a relatively simple classification task in these structured documents. However, detecting relations between arguments in the texts required more formal machinery in the guise of a context free grammar (CFG) containing rules that can be used for parsing the text. In this CFG, indicator words are used to determine, for example, conflict (however, although) or support (therefore, moreover).

Mochales Palau and Moens present pioneering research on the subject of determining the argumentative structure of natural language texts, and they show that relatively simple argumentative structures are sufficient for their purposes. Furthermore, they discuss how legal texts provide an ideal middle ground for testing and training purposes; such texts have a certain basic structure and a limited vocabulary (as opposed to, for example, arguments in newspapers, on the internet or in political debates), whilst at the same time not being so highly structured and abstract that the extraction task becomes trivial (as is the case with, for example, mathematical proofs). It is now up to the AI and Law community to build on this work by improving the argumentation mining techniques and, perhaps even more importantly, devise novel and useful ways in which the question-answering systems based on these techniques can interface with legal data.

14 Pittsburgh 2011

The Thirteenth ICAIL was held at the University of Pittsburgh from June 6th to 10th 2011 with Kevin Ashley as Conference Chair and Tom van Engers as Programme Chair. The Programme Committee was greatly expanded to over fifty members with a view to bringing in people from outside the AI and Law community, but with expertise in related technologies such as agents, normative systems and ontologies. Four papers have been selected. One (14.3) concerns a perennial issue of AI and Law, namely legal proof, while the others represent less traditional issues drawing from general AI: Bayesian reasoning (14.1), risk analysis (14.2) and agent-oriented software engineering (14.4).

14.1 Jeroen Keppens. On extracting arguments from Bayesian network representations of evidential reasoning [149]. *Commentary by Floris Bex.*

Keppens discusses two influential approaches to evidential reasoning in the law, namely structured argumentation and Bayesian networks, and shows how simple argument graphs can be extracted from the latter. He then argues that argument graphs can be used as a natural and intuitive 'interface' to the complex but powerful Bayesian network formalism. Thus, Keppens makes an important first step in unifying the probabilistic (causal) and argumentative (reason-based) approaches to evidential reasoning.

Evidential reasoning forms an important part of reasoning in the legal process [12]: the available evidence constrains which facts can be proven, which in turn are the grounds for legal consequences on the basis of the applicable legal rules or principles. Wigmore's comprehensive work on legal evidence [280] presents what can be considered the first 'formal' theory of reasoning with evidence in the law. His chart method is very similar to the argument diagramming we are familiar with in philosophy and critical thinking [94]. Furthermore, his ideas on the strength of evidence and the way in which these strengths can be used to calculate the total support for a conclusion are very similar in flavour to the way in which likelihood is computed in Bayesian networks

The argument-based approach to evidential reasoning not only captures and visualises individual inferences based on evidence and the ways in which they can be attacked, but also provides ways to compare sets of arguments based on evidence in a mathematically grounded way. The approach draws from both computational models of structured and abstract argumentation and ideas from argument diagramming and visualisation. For example, [48] shows how the Wigmore charts as presented in [145] can be rendered in the Araucaria tool for argument visualisation [221] as well as in a formal argumentation framework based on [209].

The Bayesian approach to evidential reasoning uses Bayesian Belief Networks (BBNs) to quantify the probabilistic strength of support of evidence for alternative hypotheses. It builds on Bayesian approaches to evidential reasoning in legal theory (e.g. [90]) as well as AI applications of BBNs. These BBNs have been used variously to, for example, visualise dependencies between evidence and facts in a case [131] or determine the effect of a change in the set of evidence on one's conclusions by means of sensitivity analysis [151].

Thus, as argued in the previous two paragraphs, arguments (and argument graphs) and BBNs serve different purposes in evidential reasoning. Since legal proof is reasoning under uncertainty, the mathematical underpinnings of probabilistic BBNs can provide the necessary foundations. At the same time, there is ample evidence that if we consider the way in which investigators and decision makers actually think and reason about a case, arguments are a natural component of this reasoning ([12], [47]). Therefore, Keppens argues, if we were to find a way to translate between BBNs and argument graphs, this would allow us to harness the full power of Bayesian networks whilst at the same time allowing people with relatively little mathematical training (i.e. police investigators, lawyers) to use them in practice.

Keppens presents an investigation into the similarities, differences and synergies between BBNs and argument diagrams of the kind found in [48], [94] and [221]. Like [47], Keppens distinguishes evidential reasoning of the form ‘e is evidence for c’ and causal reasoning of the form ‘c causes e’ and shows that the two are, *ceteris paribus*, symmetrical (e.g. fire causes smoke - smoke is evidence for fire). As is usual in the literature, e.g [47] and [145], evidential reasoning is here tied to argumentation (consider, e.g., the argumentation scheme for expert opinion: expert E asserts that A is evidence for A) and causal reasoning to BBNs (consider, e.g., the causal link ‘blood traces transferred from victim to suspect (T) causes blood trace on suspect matches victim (M)’ in a BBN with the associated conditional probability $P(T \mid M)$, that is, the probability that T given M). The question Keppens then asks is: *how can we extract argument diagrams from BBNs?*

Keppens presents an outline algorithm for translating BBNs to argument diagrams in a way that captures not just the main line of causal reasoning from the network in an evidential format, but also incorporates what he calls second order influences, that is, links which are not directly causal but important nonetheless, as they influence the probability of a conclusion in a BBN (e.g. the time between a burglary and the suspect’s apprehension influences the probability of there being glass shards from window of the burgled home on the suspect’s clothing).

The work by Keppens presents an exciting new frontier in the research on formal methods of evidential reasoning. It attempts to unify the ‘hard’ AI models of Bayesian networks with the more semi-formal argument diagramming approaches and thus to provide methods and tools that are both formally grounded as well as useable in forensic and legal practice. The only other work on this subject is [145] which translates Wigmore charts to Bayesian networks. Keppens’ algorithm allows for translations in the other direction. Furthermore, Keppens explores the connection between argument diagrams and BBNs in more detail by, for example, allowing the argument diagrams to incorporate second-order influences.

The connection between computational argumentation and probabilistic reasoning is an important subject for AI in general, and exactly how argumentation and Bayesian formalisms can be reconciled remains an open question. For example, Keppens does not incorporate the attack or defeat relations which are commonplace in argumentation frameworks in his algorithm. Furthermore, Keppens’ translation algorithm is currently mainly aimed at ensuring some measure of representational isomorphism, that is, ensuring that the statements and relations in the BBN are correctly translated into an argumentation diagram. Keppens makes no claims about how the results from a BBN (i.e. the posterior probabilities) relate to possible results of an argumentation framework such as ASPIC+ [208] (i.e. the acceptable propositions under some argumentation-theoretic semantics). Thus, Keppens’ algorithm does not ensure, or even aim to ensure, computational equipotence.

14.2 Mihai Surdeanu, Ramesh Nallapati, George Gregory, Joshua Walker and Christopher D. Manning. Risk Analysis for Intellectual Property Litigation [261]. *Commentary by Jack G. Conrad*

Many of the papers highlighted in this review are noteworthy for their seminal past contributions to the field of AI and Law and beyond. Many of them were published in the early to middle years of the past quarter century. By contrast, this Risk Analysis [261] paper is clearly a work that will make its impact on future research endeavors. Published in 2011, the work was circulated at a time when computational research into risk mining and risk analysis was still in its infancy. Risk analysis is a technique to identify and assess factors that may jeopardize the success of an enterprise or achieving a goal. This technique also helps to define preventive measures to reduce the probability of these factors occurring and to identify countermeasures to successfully deal with these constraints if they develop to avert possible negative effects on the competitiveness of the company [159]. Of course Wall Street analysts, insurance brokers and government regulating agencies have been performing research in this area for years, but largely in a small-scale, often manual, capacity. Thanks to advances in computational speed, memory and overall scalability, researchers today have the ability to address formerly unfathomable data mining challenges in the financial domain [153]. This work is an initial example of such an endeavor, however small its steps may be. It approaches its task by (1) focusing only on one significant sub-domain, namely, Intellectual Property (IP) cases, (2) utilizing only one strand of evidence (i.e., the prior probabilities of success while excluding the legal complexity of the subjects and entities involved), and (3) invoking a simple binary classifier to predict the outcome of the case at hand. Despite these limitations, this work is *avant garde* and should have an impact on the field well into the future. The prospects for such new forms of analysis are highly relevant and exciting.

A strength of [261] is that it defines risk in practical, legal terms - the probability of one's opponent succeeding in a legal case. Once this simple threshold is established, the authors can (and do) actively pursue models that try to use the freedom and flexibility of the features at their disposal, while admitting that they are taking small steps in framing the problem and in dealing with how much personal evidence to harness or ignore. It has served as an inspiration for members of our research group and their efforts to create Risk Mining and Risk Analysis proof of concepts and demos for the wider community. As more progress is shown in some of these areas like IP, other researchers will clearly enter the field attempting to make their mark.

The full impact of this work has yet to be felt. Given that computational Risk Analysis and Risk Mining may still be considered a fledgling field, [261] is significant as it represents an early and encouraging investigation into the prospects of using available evidence to predict key outcomes in one sub-domain of the legal field. Whether the use cases originate in Wall Street investment houses, corporate law offices, national insurance firms, or government agencies, such capabilities are clearly becoming more valued and in greater demand. At the same time, the door is left open to a host of additional analyses which may substantiate this kind of investigation, as well as extensions that may include:

(a) added sources of evidence such as the legal complexity and the entities of the case,
(b) other related content such as briefs and memoranda,
(c) materials from other sub-domains (beyond IP), such as finance, pharmacology, nuclear technology, environmental science and climatology, and
(d) harnessing more than simple binary classifiers,

to name but a few. In this sense, this research is an early attempt to analyze a subsection of a problem space. That having been accomplished, the next steps include a rich variety of advances related to those described above.

14.3 Floris Bex and Bart Verheij. Legal shifts in the process of proof [54]. *Commentary by Michał Araszkiewicz*

In this paper, Floris Bex and Bart Verheij continue the development and application of their hybrid, narrative-argumentative approach to legal reasoning advanced by the authors in their previous research. The method was most fully described and analyzed in [47]. The main context for the application of this approach they have discussed in the literature so far is evidential reasoning as they focus on criminal cases. However, in [54], the authors aim to extend their approach to cover all important aspects of legal reasoning in criminal cases, from reasoning about evidence via the discussion of legally relevant facts to the legal consequences which follow from these facts. Such a broad research perspective is rarely adopted in the field of AI and Law ([274] is one exception), although the models of reasoning from legal facts to legal consequences are very well-developed and recently there has also been a rise of interest in evidential reasoning in the field (e.g. [150]).

Bex and Verheij in [54] defend a thesis that the reasoning mechanisms in evidential reasoning and reasoning about legal consequences of facts are related and apply their narrative-argumentative approach to support this thesis by precise argumentation. Establishing the analogies between evidential reasoning and legal reasoning, they extend the theory based on their well-developed research concerning the former domain to cover the latter issue - with necessary modifications resulting from the differences concerning the content and context of these two stages of reasoning.

The essence of the hybrid approach towards evidential reasoning is that one starts with some data indicating the possibility of committing a crime. These factual data are called *explananda*. Theories which aim at explanation of these data are then constructed. In consequence, we obtain hypothetical factual stories - (possibly) coherent accounts of what might have been the case. Arguments can then be used to relate the stories with the evidence. Eventually, the stories are to be compared with each other in order to choose the most coherent story, the one which fits the evidence best.

The main contribution of [54] is to extend the hybrid approach to cover not only evidential reasoning (traditionally referred to as *quaestiones facti*) but reasoning about legal consequences of proven facts (*quaestiones iuris*) as well and to represent the relation between these two layers of legal thinking. The authors rightly draw attention to the following phenomenon: not only does the available evidence determine what (legally relevant) facts can be proven and ultimately what the legal consequences will be, but also the content of the law influences the investigative and decision making focus. This phenomenon is referred to as *legal shifts*. In consequence, based on the tool of factual story schemes developed by the authors in their previous research, they develop the concept of legal story schemes which can be seen as a holistic legal perspective on criminal cases. Bex and Verheij present a precise formal account of the relation between evidential reasoning and legal reasoning and they obtain interesting specific results related to the role of completeness of factual stories on the one hand and legal stories on the other hand. While factual stories may have evidential gaps, a legal story cannot have legal gaps at the end of the investigative process (because, in particular in criminal law, all the prerequisites of legal qualification of a given set of facts have to be satisfied in order for a legal consequence to follow). The application of legal story schemes also helps to avoid so-called legal tunnel vision, because if in a concrete case it is difficult to find evidential support corresponding to all elements of a given legal story scheme, it can make the reasoner employ another legal story scheme (a legal shift which, in turn, may influence the investigative process again).

The paper has been particularly interesting for me because it develops a view on holistic, coherentist account of legal reasoning which provides an alternative to other theories of legal coherence ([40], [120], for which see 9.3, and [11]). It must be emphasized that the project elaborated in the contribution is very ambitious: the hybrid model described here aims to cover legal reasoning in criminal cases in its entirety. Also, there are no important obstacles in the way of adapting the hybrid narrative-argumentative methodology and concept of legal story schemes developed for criminal law, to private or administrative law.

Because the paper is still relatively new, it is not a straightforward task to assess its future impact on the AI and Law community. However, the line of research is continued and recently also linked to Case-based Reasoning in the joint paper of the two authors with Trevor Bench-Capon [55]. The contribution discussed here seems to be very promising, covering all important aspects of judicial reasoning and combining the holistic coherentist perspective with formalism of argumentative approach. The project developed in the paper inspires the reader to rethink the most fundamental issues concerning legal-conceptual perspective on reality, concerning, *inter alia*, the relation between ‘real world facts’ and ‘legally qualified facts’ and the issues of justification and discovery in legal reasoning.

14.4 Alexander Boer and Tom M. van Engers. An Agent-based Legal Knowledge Acquisition Methodology for Agile Public Administration [63]. *Commentary by Erich Schweighofer*

Alexander Boer and Tom van Engers starting from their academic background in computer science and its formalisations provide a fresh view on legal knowledge acquisition with a much broader approach. The focus shifts from a reduced representation of the law-making and law-implementing process to the goal of a full representation and analysis of legally relevant processes. This is much more ambitious and legal theory has not yet solved the underlying questions. Here, computer science helps to support an in-depth analysis of legal practice.

In conceptual information retrieval, the overwhelming importance of texts, interpretations, legal authorities, legal procedures, etc. and representation and search is obvious. Legal materials have to be identified, collected, searched and analysed. The focus in the model of Alexander and Tom is on persons, agents, roles and their behaviour in a particular process. Normally, in legal knowledge acquisition methodology, the complex participation and deliberation processes are excluded. The same is true for the implementation phase. Interpretation and application is left to the user. Only final results - legal documents - count. Parliaments, courts, administrations, legal writers etc. are reduced to very special agents, the relevant authorities, allowed to promulgate rules for a particular jurisdiction. In contrast to that, Alexander and Tom want to represent and analyse the whole process. For moving on in legal theory and legal informatics, this approach can be very helpful. In the end, systems have to be assessed according to whether the results as such are of some importance for the legal practitioner.

In Alexander’s and Tom’s system, agents and their roles are the focus of research in the chosen sector of e-government. Legislative drafters, product developers, and service providers should co-operate for learning and prioritizing. The desired knowledge is not the law as such but the *effects* of the law and the *behaviours* of the agents in a modelled domain. Thus, the knowledge is much broader with no particular focus on legal rules. So called *Serious Gaming* can be used for theory construction. Multi-agent system (MAS) simulation is used as an aid for understanding the domain and the relevant agent roles in it.

Thus, Alexander and Tom move from conceptual representation of a well defined text corpora to the modelling of design activities in public administration. Law plays a major, but certainly not the only important, role. Organisational structures, people, tasks, citizens etc. determine very much the process of fulfilling the particular aim of the administration. Business process specifications, different internalisations of a legal rule in different agent roles, aims, knowledge contexts, social contexts etc. are explored. Knowledge about the text corpora is extended by compliance monitoring problems, e.g. noncompliance storylines and evidence trails, value and cost of information, preselected cases for compliance monitoring, statistical control groups, missing information etc. Lawyers call this set of information *administrative practice*, accepting a similar importance in practice but do not deal very much with this knowledge. As system designers, Alexander and Tom cannot rely on the legal text as such, but provide a solution to an ill-defined problem, which perhaps will require amendments to the law at a later stage.

Cycles of case handling, development, and legislation problem solving are proposed. The focus of formalisation is monitoring against three different conceptualisations: the organisation and its environment as a normative system, as an ontologically coherent institution and as a goal directed agent. This Multi-Agent Simulation (MAS) of agent roles is carried out using the description logic OWL, the representative MAS language Jason AgentSpeak and the very expressive rule language RIF. The example of the abstract seller agent role shows the potential for this approach. The main advantage seems to be the option of playing games for knowledge acquisition, e.g. learning about possible behaviour of agents in different administration settings and environments. As the project is very much at a model stage, however, no realistically sized examples have been presented.

I learned a lot in studying this paper. In conceptual information retrieval, we are often too focused on a textual (or multimedia) representation of the law and forget about the long-lasting and painful processes of participation, deliberation and implementation of the law. Here, a broader consideration than textual representation is needed. The aim of knowledge acquisition is extended to system design, e.g. games and simulations become important as now practice is available.

It will take some time in moving from games and simulations to a real application in practice. Many agents, different contexts and environments etc. have to be formalised and that seems to be possible only for small applications. Scaling-up remains also here the strongest barrier for a successful application. Research on knowledge acquisition from text (and audio or video) can be supportive as behaviour of agents can be more easily represented. In the future, the representation of law should not be limited to an ontological representation of textual knowledge of rules but give a comprehensive view, with proper lifting of the veil on the creation and application of legal knowledge.

15 Looking to the Future

Looking back over twenty five years of AI and Law enables us to see a great deal of development both in techniques and understanding, and in the technology and the role played by AI and Law. With respect to the latter, the changes since 1987 could not have been imagined (by me at least) at the first ICAI. The development of the World Wide Web, the enormous reduction in the cost of data storage and the enormous increase in computational power have combined to change the nature of AI and Law applications completely, both in availability and scope. On the technical side, various relationships between cases and statutes and rules, between legal knowledge and common sense knowledge, and between formal and informal approaches have provided a consistent source of inspiration and definite progress has been made in understanding these relationships better.

The story of AI and Law and ICAIL is not finished: ICAIL 2013 is planned for Rome. The concluding remarks will be given over to the Programme Chair of that conference who will offer some general reflections on the field and on the relationship between Law and AI in general and AI and Law in specific.

15.1 Towards ICAIL 2013 in Rome: the start of the next 25 years of the research program AI and Law: *Bart Verheij*

The first ICAIL of the next 25 years, the fourteenth in its existence, is planned to be held in Rome, June 10-14, 2013. Where ICAIL started its journey in Boston in 1987 in one of the capitals of the new world, we will continue our journey in a capital of the old world. In this way, we have traveled from the archetypal modern and optimistic country where the early methods of AI were created in the 1950s, and now return to the origins of the influential classic roots of the methods of law, as they were developed in ancient Rome.

It is not a coincidence that the fields of AI and Law have crossed paths, as the two fields share method and subject matter. As method, both AI and Law show the powers of what may be called *semi-formal modeling*. Where the semi-formal models of law take for instance the form of binding precedents and statutory rules, those of AI range from logical representations to robot vehicles visiting Mars. Both AI and Law know that modeling can never be purely formal nor purely informal. Modeling is always a task of finding the right balance between the order of the formal and the chaos of the informal. In law, rules have exceptions, reasons are weighed, and principles are guiding. In AI, reasoning is uncertain, knowledge is context-dependent, and behavior is adaptive.

This interest in the necessary balancing of order and chaos that is at the heart of both AI and Law points to the common subject matter that underlies the two fields: the *coordination of human behavior*. In AI, such coordination is steered by the elusive tool of intelligence, and, in law, the equally intangible technique of the rule of law is the primary coordination device. Where AI focuses for instance on the roles of knowledge, reasoning, action and interaction in coordination, the law addresses how contracts, punishment, compensation and authorities can guide human society and its inhabitants in doing the right thing.

By their shared method and subject matter, both AI and Law can be regarded as developing a *science of hermeneutic pragmatics*, which to many outside AI or law — and perhaps even to many within these fields — will sound like a *contradictio in terminis*. We, in the field of AI and Law, know that it is not. Each element in the term has to be there. 'Pragmatics' reflects the concrete goal of behavior coordination, which requires the understanding and interpretation covered by the term 'hermeneutic', of which the notoriety — partly deserved and partly undeserved — is tempered by the emphasis on 'science'. Also both AI and Law are *engineering sciences*, stressing the need to not only develop new understanding, but also build new things hands-on; whether new law or new artifices.

As a thoroughly interdisciplinary field, AI and Law is in the unique position to integrate insights from what in the Netherlands are commonly referred to as the alpha, beta and gamma sides of the sciences, roughly corresponding to the humanities, the empirical

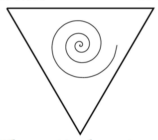

Fig. 3 Synergy between the kinds of systems investigated

sciences, and the social sciences, respectively. Also by the nature of the field, AI and Law benefits from the synergy between the different kinds of systems investigated: *theoretical systems*, such as mathematics and legal theory, are used to learn about *artificial systems*, such as software and statutes, while remaining grounded by the perspective on *natural systems*, such as human intelligence and the practices of law. (Cf. Figure 3)

Reading between the lines in this issue celebrating the first 25 years of AI and Law's main conference, it is obvious how stimulating it is to work in the field. The problems are hard, they are important, and they are far from solved. I believe — and I am not alone — that a better understanding of AI's problems can benefit law, and that a better understanding of law's problems can benefit AI. What better place than Rome could have been chosen to emphasise the promises of bringing together the more than two-and-a-half-millennia of expertise in law with the lessons of AI's half century of existence? Let us meet in Rome to extend what is possible. Perhaps not all roads lead there, but sufficiently many do.

References

1. Leonard Adelman, James Gualtieri and Sharon L. Riedl. A Multifaceted Approach to Evaluating Expert Systems. *Artificial Intelligence Design, Analysis and Manufacturing*, 8(4), 289–306, 1994.
2. Thomas Ågotnes, Wiebe van der Hoek, Moshe Tennenholtz and Michael Wooldridge. Power in normative systems. In *Proceedings of AAMAS 2009*, 145–152, 2009.
3. Khalid Al-Kofahi, Brian Grom and Peter Jackson. Anaphora resolution in the extraction of treatment history language from court opinions by partial parsing. In *Proceedings of the Seventh International Conference on AI and Law*, ACM Press, New York, 138–146. 1999.
4. Khalid Al-Kofahi, Alex Tyrrell, Arun Vachher and Peter Jackson. A machine learning approach to prior case retrieval. In *Proceedings of the Eighth International Conference on AI and Law*, ACM Press, New York, 88–93. 2001.
5. Vincent Aleven. *Teaching Case-based Argumentation Through an Example and Models*. PhD Thesis, University of Pittsburgh, Pittsburgh, PA, USA, 1997.
6. Vincent Aleven and Kevin D. Ashley. Evaluating a Learning Environment for Case-Based Argumentation Skills. In *Proceedings of the Sixth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 170–179. 1997.
7. Robert Alexy. *Theorie der Grundrechte*. Suhrkamp Verlag, Frankfurt am Main. 1985.
8. Robert Alexy. *Theorie der juristischen Argumentation. Die Theorie des rationalen Diskurses als eine Theorie der juristischen Begründung*. Suhrkamp Verlag, Frankfurt am Main, 1978.
9. Layman E. Allen and Charles S. Saxon. Better Language, Better Thought, Better Communication: The A-Hohfeld Language for Legal Analysis. In *Proceedings of the Fifth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 219–228. 1995.
10. Layman E. Allen and Charles S. Saxon. Achieving Fluency in Modernized and Formalized Hohfeld: Puzzles and Games for the LEGAL RELATIONS Language. In *Proceedings of the Sixth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 19–28. 1997.
11. A. Amaya. Formal models of coherence and legal epistemology. *Artificial Intelligence and Law* 15(4), 429–447. 2000.
12. T.J., Anderson, D.A. Schum, and W. L. Twining. *Analysis of Evidence.* 2nd edition, Cambridge University Press, Cambridge. 2005.

13. Grigoris Antoniou. Defeasible logic with dynamic priorities. *International Journal of Intelligent Systems*, 19(5), 463–472. 2004.
14. Michał Araszkiewicz. Balancing of Legal Principles and Constraint Satisfaction. In R. Winkels (ed) *Legal Knowledge and Information Systems: Jurix 2010* IOS Press, Amsterdam. 7-16. 2010.
15. Kevin D. Ashley. Toward a Computational Theory of Arguing with Precedents. In *Proceedings of the Second International Conference on Artificial Intelligence and Law*, ACM Press, New York, 93–102. 1989.
16. Kevin D. Ashley. *Modeling legal argument: Reasoning with cases and hypotheticals*. MIT Press, 1990.
17. Kevin, D. Ashley. Ontological requirements for analogical, teleological, and hypothetical legal reasoning In *Proceedings of the Twelfth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 1-10. 2009.
18. Kevin D. Ashley and Stefanie Brninghaus. A Predictive Role for Intermediate Legal Concepts. In D. Bourcier (ed) *Proceedings of Jurix 2003*, Amsterdam: IOS Press, 153-162. 2003.
19. Kevin D. Ashley and Stefanie Brninghaus.. Automatically classifying case texts and predicting outcomes. *Artificial Intelligence and Law*, 17(2) 125–165. 2009.
20. Katie Atkinson and Trevor J.M. Bench-Capon. Legal Case-based Reasoning as Practical Reasoning. *Artificial Intelligence and Law*, 13(1), 93–131. 2005.
21. Katie Atkinson, Trevor J.M. Bench-Capon, and Peter McBurney. Computational representation of persuasive argument. *Synthese*, 152, 157–206. 2006.
22. Katie Atkinson and Trevor J.M. Bench-Capon.. Practical reasoning as presumptive argumentation using action based alternating transition systems. *Artificial Intelligence*, 171(10-15), 855–874. 2007.
23. Katie Atkinson and Trevor J.M. Bench-Capon. Argumentation and standards of proof. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Law*, ACM Press, New York, 107–116. 2007.
24. Katie Atkinson, Trevor J.M. Bench-Capon, Dan Cartwright and Adam Z. Wyner. Semantic models for policy deliberation. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Law*, ACM Press. New York, 81–90. 2011.
25. Jason R. Baron and Paul Thompson. The search problem posed by large heterogeneous data sets in litigation: possible future approaches to research. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Law*, ACM Press, New York, 141–147. 2007.
26. Richard K. Belew. A Connectionist Approach to Conceptual Information Retrieval. In *Proceedings of the First International Conference on Artificial Intelligence and Law*, ACM Press, New York, 116–126. 1987.
27. Trevor J. M. Bench-Capon. Deep Models, Normative Reasoning and Legal Expert Systems. In *Proceedings of the Second international conference on Artificial Intelligence and Law*, ACM Press, New York, 37–45. 1989.
28. Trevor J.M. Bench-Capon. Neural Networks and Open Texture. In *Proceedings of the Fourth International Conference on AI and Law*, ACM Press, New York, 292–297. 1993.
29. Trevor J.M. Bench-Capon. Specification and implementation of Toulmin dialogue game. In J. Hage *et al* (eds) *Legal Knowledge-Based Systems. Jurix 1998*, Gerard Noodt Instituut, Nijmegen, 5-19. 1998.
30. Trevor J.M. Bench-Capon. The missing link revisited: The role of teleology in representing legal argument. *Artificial Intelligence and Law*, 10(1-3), 79–94. 2002.
31. Trevor J.M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3), 429–448. 2003.
32. Trevor J.M. Bench-Capon. Relating Values in a Series of Supreme Court Decisions. In: K. Atkinson (ed) *Legal Knowledge-Based Systems. Jurix 2011*. IOS Press. Amsterdam, 13-22. 2012.
33. Trevor J.M. Bench-Capon, Gwen O. Robinson, Tom W.Routen, and Marek J. Sergot. Logic programming for large scale applications in law: A formalisation of supplementary benefit legislation. In *Proceedings of the First International Conference on Artificial Intelligence and Law*, ACM Press, New York, 190–198. 1987.
34. Trevor J.M. Bench-Capon, Paul E. Dunne, and Paul H. Leng. A Dialogue Game for Dialectical Interaction with Expert Systems. In J C Rault, editor, *Proceedings of the Twelfth Annual Conference on Expert Systems and their Applications (vol. 1)*, Nanterre, 105–113. 1992.
35. Trevor J. M. Bench-Capon and Frans P. Coenen. Isomorphism and legal knowledge based systems. *Artificial Intelligence and Law*, 1(1), 65–86, 1992.
36. Trevor J. M. Bench-Capon and Geof Staniford. PLAID - Proactive Legal Assistance. *Proceedings of the Fifth International Conference on AI and Law*, ACM Press, New York. 81-88. 1995.
37. Trevor J. M. Bench-Capon and Pepijn R. S. Visser. Ontologies in legal information systems; the need for explicit specifications of domain conceptualizations. In: *Proceedings of the Sixth International Conference on Artificial Intelligence and Law*. ACM Press, New York, 132–141. 1997.
38. Trevor J. M. Bench-Capon and Giovanni Sartor. Using values and theories to resolve disagreement in law. In: J. Breuker, R. Leenes and R. Winkels (eds) *Legal Knowledge and Information Systems: Jurix 2000* IOS Press, Amsterdam, 73-84. 2000.

39. Trevor J. M. Bench-Capon and Edwina L. Rissland. Back to the Future: Dimensions Revisited. In: A. Lodder, R. Loui and A. Muntjewerff (eds) *Legal Knowledge and Information Systems: Jurix 2001* IOS Press, Amsterdam, 41-52. 2001.
40. Trevor J. M. Bench-Capon and Giovanni Sartor. A Quantitative Approach to Theory Coherence. In: A. Lodder, R. Loui and A. Muntjewerff (eds) *Legal Knowledge and Information Systems: Jurix 2001* IOS Press, Amsterdam, 53–62. 2001.
41. Trevor J. M. Bench-Capon and Giovanni Sartor. A model of legal reasoning with cases incorporating theories and values. *Artificial Intelligence*, 150(1-2) 97–143. 2003.
42. Trevor J. M. Bench-Capon and Henry Prakken. Justifying actions by accruing arguments. In P.E. Dunne and T.J.M. Bench-Capon, editors, *Computational Models of Argument. Proceedings of COMMA 2006*, IOS Press, Amsterdam, 247–258. 2006.
43. Trevor J. M. Bench-Capon and Thomas F. Gordon. Isomorphism and argumentation. In *Proceedings of the Twelfth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 11–20. 2009.
44. Trevor J. M. Bench-Capon, Henry Prakken and Wietske Visser. Argument schemes for two-phase democratic deliberation. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 21–30. 2011.
45. Donald H. Berman and Carole D. Hafner, Obstacles to the Development of Logic-Based Models of legal Reasoning. In C. Walter (ed): *Computer Power and Legal Language: The Use of Computational Linguistics, Artificial Intelligence, and Expert Systems in the Law*. Quorum Books, Westport CT., 183–214. 1988.
46. Donald H. Berman, Carole D. Hafner. Representing Teleological Structure in Case-based Legal Reasoning: The Missing Link. In: *Proceedings of the Fourth International Conference on Artificial Intelligence and Law, Amsterdam*. ACM Press, New York, 50-59. 1993.
47. Floris J. Bex. *Arguments, Stories and Criminal Evidence: A Formal Hybrid Theory.* Springer, Dordrecht. 2011.
48. Floris J. Bex, Henry Prakken, Chris Reed and Douglas N. Walton. Towards a formal account of reasoning about evidence: argumentation schemes and generalisations. *Artificial Intelligence and Law*. 11(2-3), 125–165. 2003.
49. Floris J. Bex and Henry Prakken. Reinterpreting arguments in dialogue: an application to evidential reasoning. In T.F. Gordon (ed) *Legal Knowledge and Information Systems. Jurix 2004*, IOS Press, Amsterdam, 119–129. 2004
50. Floris J. Bex, Henry Prakken and Bart Verheij. 'Formalising Argumentative Story-based Analysis of Evidence.' *Proceedings of the Eleventh International Conference on Artificial Intelligence and Law*, ACM Press, New York, 1–10. 2007.
51. Floris J. Bex, Susan W. van den Braak, Herre van Oostendorp, Henry Prakken, Bart Verheij, and Gerard A. W. Vreeswijk. Sense-making software for crime investigation: how to combine stories and arguments? *Law, Probability and Risk*, 6(1-4) 145-168. 2007
52. Floris J. Bex, Peter J. van Koppen, Henry Prakken and Bart Verheij. A Hybrid Formal Theory of Arguments, Stories and Criminal Evidence. *Artificial Intelligence and Law*, 18(2), 123-152. 2010.
53. Floris J. Bex and Henry Prakken. Investigating Stories in a Formal Dialogue Game. In P. Besnard, S. Doutre and A. Hunter (eds.), *Computational Models of Argument: Proceedings of COMMA 2008*. IOS Press,Amsterdam. 73-84. 2010.
54. Floris J. Bex and Bart Verheij. Legal shifts in the process of proof. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 11-20. 2011.
55. Floris J. Bex, Trevor J. M. Bench-Capon and Bart Verheij. What Makes a Story Plausible? The Need for Precedents. In K. Atkinson (ed) *Legal Knowledge and Information Systems. Jurix 2011: The Twenty Fourth Annual Conference.*, IOS Press, Amsterdam, 23–32. 2011
56. Jon Bing and Trygee Harwold. *Legal Decisions and Information Systems*. Universitetsforlaget, Oslo. 1977.
57. Jon Bing (ed). *Handbook of Legal Information Retrieval*. North-Holland, Amsterdam - New York - Oxford. 1984.
58. Jon Bing. Designing Text Retrieval Systems for Conceptual Searching. In *ICAIL '87: Proceedings of the First International Conference on Artificial Intelligence and Law*, ACM Press, New York, 43–51. 1987.
59. Jon Bing. Performance of Legal Text Retrieval Systems: The Curse of Boole. *Law Library Journal*, 79, 187–202, 1987.
60. David C. Blair, and M.E. Maron, An Evaluation of Retrieval Effectiveness for a Full-text Document-retrieval System. *Communications of the ACM*, 28(3), 289–299. 1985.
61. Barry W. Boehm, J.R. Brown, H. Kaspar, M. Lipow, G. McLeod, and M. Merritt. *Characteristics of Software Quality*. TRW Software Series, New Holland, NY. 1978.
62. Alexander Boer, Tom M. van Engers, and Radboud Winkels. Using Ontologies for Comparing and Harmonizing Legislation. In: *In Proceedings of the Ninth International Conference on Artificial Intelligence and Law)*. ACM Pres, New York, 161–172. 2003.

63. Alexander Boer and Tom M. van Engers. An Agent-based Legal Knowledge Acquisition Methodology for Agile Public Administration. In: *In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Law*. ACM Press, New York, 171–180. 2011
64. Susan W. Van den Braak. *Sensemaking Software for Crime Analysis*. Doctoral dissertation Department of Information and Computing Sciences, Utrecht University, 2010.
65. L. Karl Branting. Reasoning with portions of precedents. In *Proceedings of the Third International Conference on Artificial Intelligence and Law*. ACM Press, New York, 145–154. 1991.
66. L. Karl Branting. Building Explanations from Rules and Structured Cases. *International Journal of Man-Machine Studies*, 34(6):797–837, 1991.
67. L. Karl Branting. A Reduction-Graph Model of Ratio Decidendi. In *Proceedings of the Fourth international conference on Artificial Intelligence and Law*. ACM Press, New York, 40–49. 1993.
68. L. Karl Branting. A computational model of ratio decidendi. *Artificial Intelligence and Law*, 2(1), 1–31. 1993.
69. L. Karl Branting. A reduction-graph model of precedent in legal analysis. *Artificial Intelligence*, 150(1-2), 59–95. 2003.
70. Joost Breuker and Nienke den Haan. Separating World and Regulation Knowledge: Where is the Logic In: *In Proceedings of the Third International Conference on Artificial Intelligence and Law*. ACM Press, New York, 92–97. 1991
71. Joost Breuker and R. Hoekstra. Epistemology and ontology in core ontologies: FOLaw and LRICore, two core ontologies for law. In: *Proceedings of EKAW Workshop on Core ontologies*. CEUR. 2004
72. Stefanie Brüninghaus and Kevin D. Ashley. A Predictive Role for Intermediate Legal Concepts In: *In Proceedings of the Ninth International Conference on Artificial Intelligence and Law*. ACM Press, NewYork, 233-242. 2003.
73. Tom Bylander and B. Chandrasekaran. Generic tasks for knowledge-based reasoning: the “right” level of abstraction for knowledge acquisition. *International Journal on Man-Mach. Stud.* 26(2), 231–243. 1987
74. Dan Cartwright and Katie Atkinson. Using computational argumentation to support e-participation. *IEEE Intelligent Systems*, 24(5):42–52, 2009.
75. A. James Casner and W. Barton Leach. *Cases and Text on Property*. Little Brown, 1964.
76. B. Chandrasekaran. 1986, Generic tasks in knowledge-based reasoning: high-level building blocks for expert system design. *IEEE Expert* 1(3), 23–30. 1986.
77. Roderick Chisholm. Contrary-to-duty imperative and deontic logic. *Analysis*, 24(2), 33–36. 1963.
78. Alison Chorley and Trevor J.M. Bench-Capon. AGATHA: Using heuristic search to automate the construction of case law theories. *Artificial Intelligence and Law*, 13(1) 9–51, 2005.
79. William Clancey. *The epistemology of a rule-based expert system: a framework for explanation*. Technical Report STAN-CS-81-896, Stanford University, Department of Computer Science. 1981
80. Jack G. Conrad, and Daniel P. Dabney. A cognitive approach to judicial opinion structure: applying domain expertise to component analysis. *Proceedings of the Eighth International Conference on AI and Law*. ACM Press, New York, 1-11. 2001.
81. Jack G. Conrad and Frank Schilder. Opinion mining in legal blogs. In *Proceedings of the Eleventh International Conference on AI and Law*, ACM Press, New York, 231–236. 2007.
82. Jack G. Conrad, Jochen L. Leidner, Frank Schilder and Ravi Kondadadi Query-based opinion summarization for legal blog entries. In *Proceedings of the Twelfth International Conference on AI and Law*, ACM Press, New York, 167–176. 2009.
83. Rupert Cross. *Precedent in English Law, 3rd Edition*. Oxford University Press, 1979.
84. Surendra Dayal and Peter Johnson. A web-based revolution in Australian public administration. In *Proceedings of Law via the Internet*, reprinted in *Journal of Information, Law, and Technolog*y 1. 2000
85. Richard Duda, John Gasching, and Peter Hart. Model design in the PROSPECTOR consultant system for mineral exploration. In D. Michie, editor, *Expert Systems in the Micro-Electronic Age*, pages 153–167. Edinburgh University Press, Edinburgh, 1979.
86. Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–357, 1995.
87. Frans H. Van Eemeren and Rob Grootendorst. *A Systematic Theory of Argumentation. The pragma-dialectic approach.* Cambridge University Press, 2004.
88. Arthur M. Farley and Kathleen Freeman. Burden of Proof in Legal Argumentation. In *Proceedings of the Fifth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 156–164. 1995.
89. Atefeh Farzindar and Guy Lapalme LetSum, an automatic legal text summarizing system. In A. Lodder, R. Loui and A. Muntjewerff (eds) *Proceedings of the Seventeenth Annual Conference on Legal Knowledge and Information Systems, Jurix 2004*, IOS Press, Amsterdam. pages 11–18. 2004.
90. Michael O. Finkelstein and Bruce Levin. On the Probative Value of Evidence from a Screening Search. em Jurimetrics. 43, 265-90. 2003.

91. Enrico Francesconi and Ginevra Peruginelli. Integrated access to legal literature through automated semantic classification.. *Artificial Intelligence and Law*, 17(1), 31-49.. 2008.
92. Kathleen Freeman. *Towards Formalizing Dialectical Argumentation*. Phd thesis, University of Oregon, 1993.
93. Kathleen Freeman and Arthur M. Farley. A Model of Argumentation and Its Application to Legal Reasoning. *Artificial Intelligence and Law*, 4(3-4), 163–197. 1996
94. James B. Freeman. *Dialectics and the Macrostructure of Arguments: A Theory of Argument Structure.* Foris Publications, Berlin 1991.
95. Filippo Galgani and Achim Hoffman LEXA: Towards automatic legal citation classification. In *Proceedings of the Twenty-Third Australian Joint Conference on Artificial Intelligence*, 445–454. 2010.
96. M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, New York. 1979.
97. Tim van Gelder. Enhancing deliberation through computer supported argument visualization. In P.A. Kirschner, S.J.B. Shum, and C.S. Carr, editors, *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*, Springer, Berlin, 97–115. 2003.
98. Arthur Goodhart. Determining the ratio decidendi of a case. *Yale Law Journal*, 40(2), 161–183. 1930.
99. Thomas F. Gordon. Oblog-2: A Hybrid Knowledge Representation System for Defeasible Reasoning. In: *In Proceedings of the First International Conference on Artificial Intelligence and Law*, ACM Press, New York, 231–239. 1987.
100. Thomas F. Gordon. *The Pleadings Game; An Artificial Intelligence Model of Procedural Justice*. Ph.d., Technical University of Darmstadt, 1993.
101. Thomas F. Gordon. The Pleadings Game: Formalizing Procedural Justice. In: *Proceedings of the Fourth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 10–19. 1993.
102. Thomas F. Gordon. *The Pleadings Game. An Artificial Intelligence Model of Procedural Justice*. Kluwer Academic Publishers, Dordrecht/Boston/London, 1995.
103. Thomas F. Gordon and Nikos Karacapilidis. The Zeno Argumentation Framework. In *Proceedings of the Sixth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 10–18. 1997.
104. Thomas F. Gordon, Henry Prakken and Douglas N. Walton. The Carneades Model of Argument and Burden of Proof. *Artificial Intelligence*, 171(10-15), 875-896. 2007.
105. Thomas F. Gordon and Douglas N. Walton. Proof Burdens and Standards. In Iyad Rahwan and Guillermo Simari, editors, *Argumentation in Artificial Intelligence*, Springer-Verlag, Berlin, 239–260. 2009.
106. Guido Governatori. Labelling ideality and subideality. In Dov M. Gabbay and Hans Jrgen Ohlbach, editors, *Practical Reasoning*, number 1085 in LNAI, Springer-Verlag, Berlin, 291–304. 1996.
107. Guido Governatori. Representing business contracts in RuleML. *International Journal of Cooperative Information Systems*, 14(2-3), 181–216, 2005.
108. Guido Governatori, Francesco Olivieri, Simone Scannapieco, and Matteo Cristani. Superiority based revision of defeasible theories. In Mike Dean, John Hall, Antonino Rotolo, and Said Tabet, editors, *RuleML*, volume 6403 of *Lecture Notes in Computer Science*, Springer, Berlin, 104–118. 2010.
109. Matthias Grabmair and Kevin D. Ashley. Facilitating case comparison using value judgments and intermediate legal concepts. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Law*. ACM Press, New York, 50-59. 2011.
110. Katie Greenwood, Trevor J.M. Bench-Capon, and Peter McBurney. Towards a computational account of persuasion in law. In *Proceedings of the Ninth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 22–31. 2003.
111. Thomas R. Gruber. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal Human-Computer Studies* 43(5-6), 907–928. 1995.
112. Uma G. Gupta. Validation and Verification of Knowledge-based Systems: a Survey. *Journal of Applied Intelligence*, 3(4), 343–363, 1993.
113. Ben Hachey and Claire Grover. Automatic legal text summarisation: experiments with summary structuring. In *Proceedings of the Tenth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 75–84. 2005.
114. Carole D. Hafner. Conceptual organization of case law knowledge bases. In *Proceedings of the First International Conference on Artificial Intelligence and Law*, ACM Press, New York, 35–42. 1987. .
115. Carole D. Hafner and Donald H. Berman. The role of context in case-based legal reasoning: teleological, temporal, and procedural. *Artificial Intelligence and Law*, 10(1-3), 19–64. 2002.
116. Jaap Hage. Monological reason-based logic: a low-level integration of rule-based reasoning and case-based reasoning. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 30–39. 1993.
117. Jaap Hage. Teleological Reasoning in Reason-Based Logic. In *Proceedings of the Fifth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 11–32. 1995

118. Jaap Hage. *Reasoning with Rules. An Essay on Legal Reasoning and Its Underlying Logic.* Kluwer Academic Publishers, Dordrecht. 1997.
119. Jaap Hage. Goal-Based theory evaluation. In: J. Breuker, R. Leenes and R. Winkels (eds) *Legal Knowledge and Information Systems: Jurix 2000* IOS Press, Amsterdam, 59-72, 2000.
120. Jaap Hage. Formalizing legal coherence. In *Proceedings of the Eighth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 22–31. 2001.
121. Jaap Hage.*Studies in legal Logic*. Springer, Berlin. 2005.
122. Jaap Hage, Georges Span, and Arno R. Lodder. A dialogical model of legal reasoning. In *C. Grutters* et al (eds) *Legal Knowledge-Based Systems: Jurix 92*, Koninklijjke Vermade, Lelystad, 135–145. 1992.
123. Jaap Hage and Bart Verheij. Reason-Based Logic: a Logic for Reasoning with Rules and Reasons. *Information and Communications Technology Law*, 3(2-3), 171–209. 1994.
124. Jaap Hage, Ronald E.Leenes, and Arno R. Lodder. Hard cases: a procedural approach. *Artificial Intelligence and Law*, 2(2), 113–166. 1993.
125. Jaap Hage and Aleksander Peczenik. Law, Morals, and Defeasibility. *Ratio Juris* 13(3), 305-325, 2000.
126. Jaap Hage and Aleksander Peczenik. Legal Internalism, In P. Chiassoni (ed.), *The legal ought. Proceedings of the IVR mid-term congress*, Genoa, 141-170. 2001
127. Jean Hall and John Zeleznikow. Acknowledging insufficiency in the evaluation of legal knowledge-based systems: Strategies towards a broad based evaluation model. In *Proceedings of the Eighth International Conference on Artificial Intelligence and Law*, ACM Press. New York, 147–50. 2001.
128. Peter Hammond. *Representation of DHSS Regulations as a Logic Program*. Department of Computing: Research report DoC, Imperial College. 1983.
129. H.L.A. Hart. *The Concept of Law.* Clarendon Press, Oxford. 1961.
130. Gertjan van Heijst. *The Role of Ontologies in Knowledge Engineering*. Ph.D. thesis, Social Science Informatics, University of Amsterdam. 1995
131. Amanda R. Hepler, A. Philip Dawid,, and Valentina Leucari. Object-oriented graphical representations of complex patterns of evidence. em Law, Probability and Risk. 6(1-4), 275 - 293. 2007.
132. Henning Herrestad. Norms and Formalization. In *Proceedings of the Third International Conference on Artificial Intelligence and Law*, ACM Press. New York, 175–184. 1991.
133. Henning Herrestad and Christen Krogh. Obligations Directed from Bearers to Counterparts. In *Proceedings of the Fifth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 210–218. 1995.
134. Keith J. Holyoak and Paul Thagard. Analogical mapping by constraint satisfaction. *Cognitive Science*, 13(3), 295–355. 1989.
135. John F. Horty. Reasons and precedent. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 41–50. 2011.
136. Hume, David, 1739. *A Treatise on Human Nature*. Available as *A Treatise of Human Nature*, edited by L. A. Selby-Bigge, 2nd ed. revised by P.H. Nidditch, Clarendon Press, Oxford. 1975.
137. Daniel Hunter. Out of their minds: Legal theory in neural networks. *Artificial Intelligence and Law*, 7(2-3), 129–151. 1999.
138. Peter Jackson, Khalid Al-Kofahi, Alex Tyrrell and Arun Vachher Information extraction from case law and retrieval of prior cases. *Artificial Intelligence*, 150(1-2), 239–290. 2003.
139. Peter Jackson and Khalid Al-Kofahi. Human expertise and artificial intelligence in legal search. In Geist, A.; Brunschwig, C.R.; Lachmeyer, F.; Schefbeck, G., (eds) *Strukturierung der Juristischen Semantik - Structuring Legal Semantics* Bern: Editions Weblaw, p. 417-427. 2011
140. Hadassa Jakobovits and Dirk Vermeir. Dialectic semantics for argumentation frameworks. In *Proceedings of the Seventh International Conference on Artificial Intelligence and Law*, ACM Press, New York, 53–62. 1999.
141. Peter Johnson and David Mead. Legislative knowledge base systems for public administration: Some practical issues. In *Proceedings of the Third International Conference on Artificial Intelligence and Law*, ACM Press, New York, 108–117. 1991.
142. Andrew J.I. Jones and Ingmar Pörn. Ideality, sub-ideality and deontic logic. *Synthese*, 65(2), 275–290. 1985.
143. Andrew J.I. Jones and Ingmar Pörn. “Ought” and “Must”. *Synthese*, 66(1), 89–93. 1986.
144. Natalia Juristo Juzgado and Jos L. Moran. Common Framework for the Evaluation Process of KBS and Conventional Software. *Knowledge Based Systems*, 1(2), 145–159. 1998.
145. Joseph B. Kadane and David A. Schum. *A Probabilistic Analysis of the Sacco and Vanzetti Evidence.* John Wiley and Sons, New York. 1996.
146. Nikos I. Karacapilidis and Dimitris Papadias. Computer supported argumentation and collaborative decision making: the HERMES system. *Information Systems*, 26(4), 259–277. 2001.

147. Nikos Karousos, Spyridon Papaloukas, Nektarios Kostaras, Michalis Nik Xenos, Manolis Tzagarakis and Nikos I. Karacapilidis. Usability evaluation of web-based collaboration support systems: The case of CoPe_it! In *Proceedings of the Third World Summit on the Knowledge Society*, volume 111 of *Communications in Computer and Information Science*, Springer, Berlin, 248–258. 2010.
148. Jurgen Karpf. Quality assurance of legal expert systems. In *Pre-Proceedings of the Third International Conference on Logica, Informatica, Diritto*, CNR, Florence, 411–440. 1989.
149. Jeroen Keppens. On extracting arguments from Bayesian network representations of evidential reasoning. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 141–150. 2011.
150. Jeroen Keppens and Burkhard Schafer Knowledge based crime scenario modelling. *Expert Systems with Applications* 30(2), 203-222, 2006.
151. Jeroen Keppens, Qiang Shen and Chris Price. Compositional Bayesian Modelling for Computation of Evidence Collection Strategies. em Applied Intelligence. 35(1), 134-161. 2011.
152. Paul A. Kirschner, Simon J. Buckingham Shum, and Chad S. Carr (editors). *Visualizing argumentation: Software tools for collaborative and educational sense-making*. Springer, Berlin. 2003.
153. Shimon Kogan, Dimitry Levin, Bryan R. Routledge, Jacob S. Sagi and Noah A. Smith. Predicting Risk from Financial Reports with Regression. In *Proceedings of the Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics, Stroudsburg, PA, 272–280. 2009.
154. Robert A. Kowalski and Francesca Toni. Abstract Argumentation. *Artificial Intelligence and Law*, 4(3-4), 275–296. 1996.
155. Marc Lauritsen. Intelligent tools for managing factual arguments. In *Proceedings of the Tenth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 95–104. 2005.
156. Ronald E. Leenes. Burden of Proof in Dialogue Games and Dutch Civil Procedure. In *Proceedings of the Eighth international conference on Artificial intelligence and law*, ACM Press, New York, 109–118. 2001.
157. Phillip Leith. ELI: An expert legislative consultant. In *Proceedings of the IEE Conference on Man/Machine Systems* UMIST Conference Publication 212. 1982.
158. Edward H. Levi. *An Introduction to Legal Reasoning.* University of Chicago Press. 1949.
159. Jochen L. Leidner and Frank Schilder. Hunting for the Black Swan: Risk Mining from Text. In *Proceedings of the Association for Computational Linguistics (ACL)*, Association for Computational Linguistics Stroudsburg, PA, 54–59. 2010.
160. Lars Lindahl and Jan Odelstad. Intermediaries and intervenients in normative systems. *Journal of Applied Logic* 6(2), 229–250. 2008
161. John W. Lloyd. *Foundations of Logic Programming, 2nd Edition*. Springer, Berlin. 1987.
162. Arno R. Lodder. *DiaLaw. On Legal Justification and Dialogical Models of Argumentation*. Law and Philosophy Library. Kluwer Academic Publishers, Dordrecht/Boston/London. 1999.
163. Arno R. Lodder and Aimée Herczog. DiaLaw A Dialogical Framework for Modeling Legal Reasoning. In *Proceedings of the Fifth International Conference on Artificial intelligence and Law*, ACM Press, New York, 146–155. 1995.
164. Paul Lorenzen and Kuno Lorenz. *Dialogische Logik*. Wissenschaftliche Buchgesellschaft, Darmstadt. 1978.
165. Ronald P. Loui. Process and policy: resource-bounded non-demonstrative reasoning. *Computational Intelligence*, 14(1), 1–38, 1998.
166. Ronald P. Loui, Jeff Norman, Jon Olson, and Andrew Merrill. A Design for Reasoning with Policies, Precedents and Rationales. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 202–211. 1993.
167. Ronald P. Loui and Jeff Norman. Rationales and argument moves. *Artificial Intelligence and Law*, 3(3), 159–189, 1995.
168. Ronald P. Loui, Jeff Norman, Joe Altepeter, Dan Pinkard, Dan Craven, Jessica Linsday, Mark A. Foltz. Progress on Room 5: a testbed for public interactive semi-formal legal argumentation. In *Proceedings of the Sixth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 207–214. 1997.
169. Rolf Luehrs, Thomas Malsch, and K. Voss. Internet, discourses and democracy. In *New Frontiers in Artificial Intelligence*, LNCS 2253, pages 67–74. Springer-Verlag, Berlin, LNCS 2253, pages 67–74. 2001.
170. C.D. Manning, P. Raghavan and H. Schtze, (Eds.). *Introduction to Information Retrieval*. Cambridge University Press. 2008.
171. Ann Macintosh, Thomas F. Gordon, and Alastair Renton. Providing argument support for e-participation. *Journal of Information Technology and Politics*, 6(1), 43–59. 2009.
172. Ejan Mackaay and Pierre Robillard. Predicting judicial decisions: The nearest neighbour rule and visual representation of case patterns. *Datenverarbeitung im Recht*, 3(3-4), 302–331. 1974.

173. J.D. Mackenzie. Question-begging in non-cumulative systems. *Journal of Philosophical Logic*, 8(1) 117–133. 1979.
174. L. Thorne McCarty and N. S. Sridharan. The Representation of an Evolving System of Legal Concepts: II. Prototypes and Deformations. In *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, William Kaufmann, MA, 46–253. 1981.
175. L. Thorne McCarty. Permissions and obligations. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, William Kaufmann, MA, 287–294, 1983.
176. L. Thorne McCarty. Intelligent legal information systems: Problems and prospects. In Colin Campbell, editor, *Data Processing and the Law*, Sweet and Maxwell, London, 125-151. 1984.
177. L. Thorne McCarty. A Language for Legal Discourse I: Basic Features. In *Proceedings of the Second International Conference on Artificial Intelligence and Law*, ACM Press, New York, 180–189. 1989.
178. L. Thorne McCarty. Invited address: On the role of prototypes in appellate legal argument. In *Proceedings of the Tird international conference on Artificial Intelligence and Law*, ACM Press, New York, 185–190. 1991.
179. L. Thorne McCarty. An Implementation of Eisner v. Macomber. In *Proceedings of the Fifth International Conference on Artificial Intelligence and Law*,ACM Press, New York, 276–286. 1995.
180. Raquel Mochales and Marie-Francine Moens, Argumentation mining: the detection, classification and structure of arguments in text In *Proceedings of the Twelfth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 98–107. 2009.
181. Raquel Mochales and Marie-Francine Moens. Argumentation mining. *Artificial Intelligence and Law*, 19(1), 1-22. 2011.
182. Sanjay Mogdil and Trevor J.M. Bench-Capon. Metalevel argumentation. *Journal of Logic and Computation* 21(6), 959-1003. 2011
183. Sanjay Modgil and Henry Prakken. Revisiting preferences and argumentation. In Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, IJCAI/AAAI, 1021–1026. 2011.
184. Marie-Francine Moens, Caroline Uyttendaeleand Jos Dumortier. Abstracting of Legal Cases: The SALOMON Experience. In *Proceedings of the Second International Conference on Artificial Intelligence and Law*, ACM Press New York, 114–122. 1997.
185. Bernard Moulin, Hengameh Irandoust, Micheline Bélanger and G. Desbordes. Explanation and argumentation capabilities: Towards the creation of more persuasive agents. *Artificial Intelligence Review*, 17(3), 169–222. 2002.
186. Nils J. Nilsson. *Principles of Artificial Intelligence*. Springer, Oxford. 1982.
187. Robert M. O'Keefe and Daniel E. O'Leary. Expert System Verification and Validation: a Survey and Tutorial. *Artificial Intelligence Review*, 7(1) 3–42. 1993.
188. Monica Palmirani, Guido Governatori and Giuseppe Contissa. Modelling temporal legal rules. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Law*, ACM Press New York, 131-135. 2011
189. Bo Pang and Lillian Lee Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), pp. 1-135. 2008.
190. Mark C. Paulk, Charles V. Weber, Bill Curtis, and Mary Beth Chrissis. *Capability Maturity Model for Software*. Addison-Wesley, Boston. 1995.
191. Nancy Pennington and Reid Hastie. The Story Model for Juror Decision Making. In R. Hastie(ed) *Inside the Juror: The Psychology of Juror Decision Making*. Cambridge University Press, Cambridge, 192-221. 1993.
192. Aleksander Peczenik. *On Law and Reason*. Springer, Berlin. 2nd Edition 2008.
193. Aleksander Peczenik and Jaap Hage. Legal Knowledge about What?. *Ratio Juris* 13(3), 325-345, 2000.
194. Chaim Perelman and Lucie Olbrechts-Tyteca. *The New Rhetoric: A Treatise on Argumentation*. University of Notre Dame Press, Notre Dame. 1969.
195. Lothar Philipps. Artificial Morality and Artificial Law. *Artificial Intelligence and Law*, 7(2), 115–128. 1999.
196. Lothar Philipps and Giovanni Sartor. Introduction: from legal theories to neural networks and fuzzy reasoning. *Artificial Intelligence and Law*, 7(2-3), 51–63, 1999.
197. Charles S. Pierce. *Collected papers of Charles Sanders Peirce.* Harvard University Press. 1931.
198. John Pollock. Defeasible Reasoning. *Cognitive Science*, 11(4), 481–518. 1987.
199. Henry Prakken. A Logical Framework for Modelling Legal Argument. *Proceedings of the Fourth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 192-201. 1993.
200. Henry Prakken. From Logic to Dialectics in Legal Argument. In: *Proceedings of the Fifth International Conference on Artificial Intelligence and Law*. ACM Press, New York, 165–74. 1995
201. Henry Prakken. An exercise in formalising teleological case-based reasoning. In: J. Beuker, R. Leened and R. Winkels (eds) *Legal Knowledge and Information Systems: Jurix 2000* IOS Press, Amsterdam, 49-57. 2000.

202. Henry Prakken. Modelling Defeasibility in Law: Logic or Procedure? *Fundamenta Informatica*, 48(2-3), 253–271. 2001.
203. Henry Prakken. Modelling reasoning about evidence in legal procedure. In *Proceedings of the Eighth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 119–128. 2001.
204. Henry Prakken. An exercise in formalising teleological case-based reasoning. *Artificial Intelligence and Law*, 10(1-3), 111–133. 2002.
205. Henry Prakken. A study of accrual of arguments, with applications to evidential reasoning. In *Proceedings of the Tenth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 85–94. 2005.
206. Henry Prakken. Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation*, 15(6), 1009–1040. 2005.
207. Henry Prakken. A formal model of adjudication dialogues. *Artificial Intelligence and Law*, 16(3), 305–328. 2008.
208. Henry Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1(2), 93–124. 2010.
209. Henry Prakken and Giovanni Sartor. A Dialectical Model of Assessing Conflicting Argument in Legal Reasoning. *Artificial Intelligence and Law*, 4(3-4), 331–368. 1996.
210. Henry Prakken and Giovanni Sartor. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-classical Logics*, 7(1), 25–75. 1997.
211. Henry Prakken and Giovanni Sartor. Reasoning with Precedents in a Dialogue Game. In *Proceedings of the Sixth International Conference on Artificial Intelligence and Law*. ACM Press, New York, 1–9. 1997.
212. Henry Prakken and Giovanni Sartor. Modelling Reasoning with Precedents in a Formal Dialogue Game. *Artificial Intelligence and Law*, 6(2-4), 231–287. 1998.
213. Henry Prakken and Giovanni Sartor. Presumptions and Burdens of Proof. In Tom M. van Engers (ed) *Legal Knowledge and Information Systems: JURIX 2006: The Nineteenth Annual Conference*, : IOS Press, Amsterdam, 21-30. 2006.
214. Henry Prakken and Giovanni Sartor. Formalising Arguments about the Burden of Persuasion. *Proceedings of the Eleventh International Conference on Artificial Intelligence and Law*. ACM Press, New York, 176-185. 2007.
215. Henry Prakken and Giovanni Sartor. More on Presumptions and Burdens of Proof. In E. Francesconi, G. Sartor and D. Tiscorina (eds) *Legal Knowledge and Information Systems. Jurix 2008: The Twenty-First Annual Conference.* IOS Press, Amsterdam, 21-30. 2008.
216. Henry Prakken and Giovanni Sartor. A Logical Analysis of Burdens of Proof. Legal Evidence and Burden of Proof. In H. Kaptein, H. Prakken and B. Verheij (eds) *Legal Evidence and Burden of Proof: Statistics, Stories, Logic*. Ashgate, Farnham, 223- 253. 2009.
217. Henry Prakken and Giovanni Sartor. On modelling burdens and standards of proof in structured argumentation. In K. Atkinson (ed) *Legal Knowledge and Information Systems (Jurix 2011)*, IOS Press, Amsterdam, 83–92. 2011.
218. Henry Prakken, Chris Reed, and Douglas N. Walton. Argumentation Schemes and Burden of Proof. In *Proceedings of the Fourth Workshop on Computational Models of Natural Argument*, ECAI, Valencia, 81–86, 2004.
219. Henry Prakken, Chris Reed, and Douglas N. Walton. Dialogues about the Burden of Proof. *Proceedings of the Tenth International Conference on AI and Law*. ACM Press, New York, 15-124. 2005
220. Susan L. Price, Marianne Lykke Nielsen, Lois M. L. Delcambre, Peter Vedsted and Jeremy Steinhauer. Using semantic components to search for domain-specfic documents: An evaluation from the system perspective. In *Information Systems*, 34(8), 724–752. 2009.
221. Chris Reed and Glenn Rowe. Araucaria: Software for Argument Analysis, Diagramming and Representation. *International Journal of AI Tools*. 13(4), 961-980. 2004.
222. Raymond Reiter. A logic for default reasoning. In *Artificial Intelligence*, 13(1-2), 81–132. 1980.
223. Nicholas Rescher. *Dialectics: A Controversy-Oriented Approach to the Theory of Knowledge*. State University of New York Press, Albany. 1977.
224. Elaine Rich and Kevin Knight. *Artificial Intelligence*. Second Edition, McGraw Hill, New York. 1991.
225. Edwina L. Rissland. Examples in Legal Reasoning: Legal Hypotheticals.. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, pages 90–93, William Kaufman, MA. 1983.
226. Edwina L. Rissland and Robert T. Collins The Law as Learning System. In *Proceedings Eighth Annual Cognitive Science Society Conference*, Amherst, MA, 500-513. 1986.
227. Edwina L. Rissland and Kevin D. Ashley. A Case-Based System for Trade Secrets Law.' *Proceedings of the First International Conference on AI and Law*. ACM Press, New York, 60-66. 1987.
228. Edwina L. Rissland and David B. Skalak. Interpreting Statutory Predicates. In *Proceedings of the Second International Conference on Artificial Intelligence and Law*, ACM Press, New York, 46–53. 1989.

229. Edwina L. Rissland and David B. Skalak. Combining Case-Based and Rule-Based Reasoning: A Heuristic Approach. In *Proceedings of Eleventh International Joint Conference on Artificial Intelligence*, Morgan Kaufmann, CA, 524–530. 1989.
230. Edwina L. Rissland and David B. Skalak. CABARET: Statutory Interpretation on a Hybrid Architecture. In *International Journal of Man Machine Studies*, 34(6), 39–887. 1991.
231. Edwina L. Rissland, David B. Skalak and M. Timur Friedman. BankXX: A Program to Generate Argument Through Case-Base Search. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 117–124. 1993.
232. Edwina L. Rissland, and Jody J. Daniels. A hybrid CBR-IR approach to legal information retrieval. In *Proceedings of the Fifth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 52–61. 1995.
233. Edwina L. Rissland and M. Timur Friedman. Detecting Change in Legal Concepts. In *Proceedings of the Fifth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 127–136. 1995.
234. Edwina L. Rissland, David B. Skalak, and M. Timur Friedman. BankXX: Supporting legal arguments through heuristic retrieval. *Artificial Intelligence and Law*, 4(1), 1–71. 1996.
235. Edwina L. Rissland, David B. Skalak, and M. Timur Friedman. Evaluating a legal argument program: The BankXX experiments. *Artificial Intelligence and Law*, 5(1-2), 1–74. 1997.
236. Edwina L. Rissland and Kevin D. Ashley. A Note on Dimensions and Factors. *Artificial Intelligence and law*, 10(1-3), 65–77. 2002.
237. Edwina L. Rissland, Kevin D. Ashley. and L. Karl Branting. Case-based reasoning and law. *The Knowledge Engineering Review*, 20(3) 293–298, 2005.
238. Edwina L. Rissland Black Swans, Gray Cygnets and Other Rare Birds. In *Case-Based Reasoning Research and Development, Eighth International Conference on Case-Based Reasoning*, Springer, Berlin, 6-13. 2009.
239. Horst W.J. Rittel and Marvin M. Webber. Dilemas in a general theory of planning. *Policy Sciences*, 4(2), 155–169, 1973.
240. Bram Roth and Bart Verheij. Dialectical arguments and case comparison. In T. Gordon (ed)*Legal Knowledge and Information Systems. Jurix 2004: The Seventeenth Annual Conference*, IOS Press, Amsterdam, 99–108. 2004.
241. Patrick Saint-Dizier. Processing natural language arguments with the TextCoop platform. *Journal of Argumentation and Computation*, 3(1), 49–82. 2012
242. Gerard M. Salton, Andrew K.C. Wong, and Chung-Shu Yang. A Vector Space Model for Automatic Indexing. *Communications of the ACM*, 18(11), 613–620. 1975.
243. Giovanni Sartor. A Simple Computational Model for Nonmonotonic and Adversarial Legal Reasoning. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Law*. ACM Press, New York, 1-9. 1993.
244. Giovanni Sartor. Teleological Arguments and Theory-Based Dialectics. *Artificial Intelligence and Law*, 10(1-3), 95–112, 2002.
245. Giovanni Sartor. Fundamental Legal Concepts: A Formal and Teleological Characterisation. *Artificial Intelligence and Law* 14(1-2), 101–142. 2006
246. Giovanni Sartor. Doing justice to rights and values: teleological reasoning and proportionality. *Artificial Intelligence and Law*, 18(2), 175–215. 2010.
247. Roger C. Schank and Robert P. Abelson. *Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures.* Lawrence Erlbaum, Hillsdale, New Jersey. 1997.
248. Oliver Scheuer, Frank Loll, Niels Pinkwart and Bruce M. McLaren. Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-Supported Collaborative Learning*, 5(1), 43–102. 2010.
249. D. A. Schlobohm and L. Thorne McCarty. EPS II: Estate Planning with Prototypes. In *Proceedings of the Second International Conference on Artificial Intelligence and Law*, ACM Press, New York, 1-10. 1989.
250. Erich Schweighofer and Werner Winiwarter. Legal Expert System KONTERM - Automatic Representation of Document Structure and Contents. In *Database and Expert Systems Applications*, Springer, Berlin, 486–497. 1993.
251. Erich Schweighofer, Andreas Rauber and Michael Dittenbach. Automatic Text Representation, Classification and Labeling in European Law. In *Proceedings of the Eighth International Conference on Artificial Intelligence and Law*, ACM Press, New York, 78–87. 2001.
252. Erich Schweighofer. *Legal Knowledge Representation, Automatic Text Analysis in Public International and European Law.* Kluwer Law International, The Hague, 1999.
253. Erich Schweighofer. Computing Law: From Legal Information Systems to Dynamic Legal Electronic Commentaries. In: Cecilia Magnusson Sjberg and Peter Wahlgren (eds.), *Festskrift till Peter Seipel.*, Norsteds Juridik AB, Stockholm, 569-588. 2006.

254. Marek J. Sergot. Prospects for representing the law as logic programs. In K.L. Clark and S.A. Tarnlund, editors, *Logic Programming*, Academic Press, London, 33–42. 1982.
255. Marek J. Sergot, Fariba Sadri, Robert A. Kowalski, Frank Kriwaczek, Peter Hammond and H. T. Cory. The British Nationality Act as a Logic Program. *Communications of the ACM*, 29(5), 370-386. 1986
256. Edward. H. Shortliffe. *Computer-based medical consultations, MYCIN*. Artificial intelligence series. Elsevier, 1976.
257. David B. Skalak and Edwina L. Rissland. Argument Moves in a Rule-Guided Domain. In *Proceedings of the Third International Conference on Artificial Intelligence and Law*. ACM Press, New York, 1–11, 1991.
258. David B. Skalak and Edwina L. Rissland. Arguments and cases: An inevitable intertwining. *Artificial Intelligence and Law*, 1(1), 3-44. 1992.
259. J.C. Smith. The Use of Lexicons in Information Retrieval in Legal Databases. In *Proceedings of the Fifth International Conference on Artificial Intelligence and Law*. ACM Press, New York, 78–87. 1997.
260. Ronald Stamper. LEGOL: Modelling legal rules by computer. In B. Niblett (ed) *Computer Science and Law*, Cambridge University Press, 45–71, 1980.
261. Mihai Surdeanu, Ramesh Nallapati, George Gregory, Joshua Walker and Christopher D. Manning. Risk Analysis for Intellectual Property Litigation. In *Proceedings of the Thirteenth International Conference on AI and Law*, ACM Press, New York, 231–236. 2011.
262. Richard Susskind. Expert Systems in Law: Out of the Research Laboratory and into the Marketplace. *Proceedings of the First International Conference on AI and Law*. ACM Press, New York, 1-8. 1987
263. Simone Teufel. *Argumentative Zoning: Information Extraction from Scientific Text*. PhD thesis, University of Edinburgh, 1999.
264. Simone Teufel and Marc Moens. Summarizing Scientific Articles – Experiments with Relevance and Rhetorical Status. *Computational Linguistics*, 28(4), 409-445. 2002.
265. Stephen E. Toulmin. *The Uses of Argument*. Cambridge University Press. 1958.
266. Alan Tyree. *Expert systems in law*. Prentice Hall, New Jersey. 1989.
267. Andre Valente and Joost Breuker. ON-LINE: An Architecture for Modelling Legal Information. In: *Proceedings of the Fifth International Conference on Artificial Intelligence and Law*. ACM Press, New York, 307–315. 1995
268. Saskia van de Ven and Joost Breuker and Rinke Hoekstra and Lars Wortel Automated Legal Assessment in OWL 2. In E. Francesconi, G. Sartor and D. Tiscorina (eds) *Proceedings of Jurix 2008*, IOS Press, Amsterdam, 170–175, 2008.
269. Bart Verheij. *Virtual arguments: on the Design of Argument Assistants for Lawyers and Other Arguers*. T.M.C. Asser Press, The Hague. 2005.
270. Bart Verheij and Floris J. Bex. Accepting the truth of a story about the facts of a criminal case. In H. Kaptein, H. Prakken and B. Verheij (eds) *Legal Evidence and Proof: Statistics, Stories, Logic*. Ashgate, Farnham, 161-193. 2009
271. Gerard Vreeswijk and Henry Prakken. Credulous and Sceptical Argument Games for Preferred Semantics. In Proceedings of JELIA 2000, Springer, Berlin, 239–253. 2000.
272. Anita de Waard and Joost Kircz. Modeling Scientific Research Articles - Shifting Perspectives and Persistent Issues. In *Proceedings of ELPUB 2008 International Conference on Electronic Publishing* Toronto, Canada. ELPUB Digital Library, 234–245. 2008.
273. William A. Wagenaar, Peter J. van Koppen and Hans F.M. Crombag. *Anchored Narratives: The Psychology of Criminal Evidence.* Harvester Wheatsheaf, Hertfordshire, England. 1993
274. Vern R. Walker. Visualizing the Dynamics around the Rule/Evidence Interface in Legal Reasoning. *Law, Probability and Risk* Vol 6(1-4), 5-22. 2007.
275. Douglas N. Walton. *Argumentation Schemes for Presumptive Reasoning*. Lawrence Erlbaum Associates, Mahwah, NJ. 1996.
276. Douglas N. Walton. *Fundamentals of Critical Argumentation*. Cambridge University Press. 2006.
277. Douglas N. Walton, Chris Reed, and Fabrizio Macagno. *Argumentation Schemes.* Cambridge University Press. 2008.
278. Mayeh Wardeh, Trevor J.M. Bench-Capon, and Frans P. Coenen. Padua: a protocol for argumentation dialogue using association rules. *Artificial Intelligence and Law*, 17(3), 183–215. 2009.
279. Rosina O. Weber, , K. D. Ashley, and Stefanie Brüninghaus. Textual case-based reasoning. *Knowledge Engineering Review*, 20(3), 255–260. 2005.
280. John H. Wigmore. *The Principles of Judicial Proof or the Process of Proof as Given by Logic, Psychology, and General Experience, and Illustrated in Judicial Trials.* Little, Brown and Company, Boston, MA. 1913.
281. Adam Z. Wyner. An ontology in OWL for legal case-based reasoning. *Artificial Intelligence and Law*, 16(4), 361–387. 2008.

282. Adam Z. Wyner, Trevor J.M. Bench-Capon, and Katie Atkinson. Arguments, values and baseballs: Representation of Popov v. Hayashi. In A. Lodder and L. Mommens (eds) *Legal Knowledge and Information Systems (Jurix 2007)*, IOS Press, Amsterdam, 151–160. 2007.
283. Adam Z. Wyner and Wim Peters. Lexical semantics and expert legal knowledge towards the identification of legal case factors. In R. Winkels (ed) *Legal Knowledge and Information Systems (Jurix 2010)*. IOS Press, Amsterdam pages 127–136. 2010.
284. Adam Z. Wyner. Towards annotating and extracting textual legal case elements. *Informatica e Diritto: Special Issue on Legal Ontologies and Artificial Intelligent Techniques*, 19(1-2), 9–18. 2010.
285. Adam Z. Wyner, Trevor J.M. Bench-Capon, and Katie Atkinson. Formalising argumentation about legal cases. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Law*. ACM Press New York, 1–10.2011.
286. Hajime Yoshino and Tokuyasu Kakuta. The Knowledge Representation of Legal Expert System LES-3.3 with Legal Metainference. *Proceedings of The Sixth International Symposium of Legal Expert System Association*. LESA, Tokyo, 1-9. 1993
287. Hajime Yoshino. The Systematization of Legal Meta-Inference. In *Proceedings of the Fifth International Conference on Artificial Intelligence and Law*. ACM Press, New York, 266-275. 1995
288. Hajime Yoshino. On the Logical Foundations of Compound Predicate Formulae for Legal Knowledge Representation. *Artificial Intelligence and Law* 5:(1-2), 77-96. 1997.
289. Hajime Yoshino. The systematization of law in terms of the validity. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Law*. ACM Press, New York, 121-25. 2011.