Time for AI and Society

PROCEEDINGS OF THE AISB'00 SYMPOSIUM ON ARTIFICIAL INTELLIGENCE AND LEGAL REASONING

17th-20th April, 2000
University of Birmingham

AISB'00 Convention

17th-20th April 2000

University of Birmingham
England

Proceedings of the AISB'00 Symposium on

Artificial Intelligence and Legal Reasoning

Published by

The Society for the Study of Artificial Intelligence and the Simulation of Behaviour

United Kingdom

http://www.cogs.susx.ac.uk/aisb/

ISBN 1 902956 11 1

Printed at the University of Birmingham, Edgbaston, Birmingham B15 2TT, England.

Contents

The AISB'00 Convention

The millennial nature of current year, and the fact that it is also the University of Birmingham's centennial year, made it timely to have the focus of this year's Convention be the question of interactions between AI and society. These interactions include not just the benefits or drawbacks of AI for society at large, but also the less obvious but increasingly examined ways in which consideration of society can contribute to AI. The latter type of contribution is most obviously on the topic of societies of intelligent artificial (and human) agents. But another aspect is the increasing feeling in many quarters that what has traditionally been regarded as cognition of a single agent is in reality partly a social phenomenon or product.

The seven symposia that largely constitute the Convention represent various ways in which society and AI can contribute to or otherwise affect each other. The topics of the symposia are as follows: Starting from Society: The Application of Social Analogies to Computational Systems; AI Planning and Intelligent Agents; Artificial Intelligence in Bioinformatics; How to Design a Functioning Mind; Creative and Cultural Aspects of AI and Cognitive Science; Artificial Intelligence and Legal Reasoning; and Artificial Intelligence, Ethics and (Quasi-)Human Rights. The Proceedings of each symposium is a separate document, published by AISB. Lists of presenters, together with abstracts, can be found at the convention website, at http://www.cs.bham.ac.uk/~mgl/aisb/.

The symposia are complemented by four plenary invited talks from internationally eminent AI researchers: Alan Bundy ("what is a proof?"- on the sociological aspects of the notion of proof); Geoffrey Hinton ("how to train a community of stochastic generative models"); Marvin Minsky ("an architecture for a society of mind"); and Aaron Sloman ("from intelligent organisms to intelligent social systems: how evolution of meta-management supports social/ cultural advances"). The abstracts for these talks can be found at the convention website.

We would like to thank all who have helped us in the organization, development and conduct of the convention, and especially: various officials at the University of Birmingham, for their efficient help with general conference organization; the Birmingham Convention and Visitor Bureau for their ready help with accommodation arrangements, including their provision of special hotel rates for all University of Birmingham events in the current year; Sammy Snow in the School of Computer Science at the university for her secretarial and event-arranging skills; technical staff in the School for help with various arrangements; several research students for their volunteered assistance; the Centre for Educational Technology and Distance Learning at the university for hosting visits by convention delegates; the symposium authors for contributing papers; the Committee of the AISB for their suggestions and guidance; Geraint Wiggins for advice based on and material relating to AISB'99; the invited speakers for the donation of their time and effort; the symposium chairs and programme committees for their hard work and inspirational ideas; the Institue for Electrical Engineers for their sponsorship; and the Engineering and Physical Sciences Research Council for a valuable grant.

John Barnden & Mark Lee

Symposium on Artificial Intelligence and Legal Reasoning

17 April 2000

Chair:
Donald Peterson

Programme Committee:
John Barnden (University of Birmingham, UK)
Trevor Bench-Capon (University of Liverpool, UK)
Indira Carr (University of Kent at Canterbury, UK)
Marcelo Dascal, (University of Tel Aviv)
Ephraim Nissan (University of Greenwich)
Peter Tillers (Yeshiva University)
Yorick Wilks (University of Sheffield)

The theme of the present Convention is "Time for AI and Society", and as we begin the new millennium it seems appropriate to examine the prospects which AI has for modelling and supporting reasoning in legal contexts. The topic is of interest to pure AI, since legal argument provides a relatively explicit and well understood domain of real-life human reasoning; and it is of interest to applied AI and legal practice, since the use of computer-based systems is increasing in this as in other professions.

The papers presented at this Symposium address a number of issues and describe a number of systems in the field, including the use of knowledge-based systems in copyright management, case-based decision support for use in criminal law, the modelling of legal reasoning with case precedents, an argumentation formalism for use in intelligent systems in regulation, an electronic judge's assistant, the issue of probability in legal narratives, the issue of conflict and vagueness of legal principles, and a model of distributed belief revision suitable for judicial contexts.

Donald Peterson.

Copyright regulation in cyber-space: the case for a knowledge based approach using argumentation

Andrew Stranieri and John Zeleznikow

Donald Berman Laboratory for Information Technology and Law,
Applied Computing Research Institute,
La Trobe University, Bundoora, Victoria, Australia, 3083
E-mail: {stranier,johnz}@cs.latrobe.edu.au
WWW: http://www.cs.latrobe.edu.au/research/dbc/dbc.html

Abstract

Arguments concerning the regulation of copyright on the internet range from one extreme calling for the abolition of existing copyright principles to another extreme that advocates minimal change. Currently, many governments are amending copyright laws to implement minimal change. However, various commentators claim that this, in addition to the development of locking software that restricts access to copyright protected works, protects authors so well that free access to works for public benefit purposes such as research and study is threatened. In this paper, we survey issues concerning copyright in cyber-space and present a framework for the deployment of knowledge based systems that aim to provide the required protection for authors without denying others access to works for public benefit. This approach regulates copyright in cyber-space using a combination of technological, legal and economic means by representing knowledge about intellectual property as arguments based on the structure proposed by the philosopher, Toulmin.

1 Introduction

Copyright law, in most jurisdictions encourage individual creativity by vesting a bundle of rights to the author of a literary, dramatic, musical or artistic work so that he/she can control the reproduction and transmission of created works. However, the need to encourage individual creativity is balanced against the need to ensure appropriate public benefit. This is realised in copyright law, by the provision of exceptions to the author's rights for purposes associated with the public good. For example, in Australia, the Copyright Act 1968, provides an author with the exclusive right to reproduce a book she has written. However, the same act lists fair use exceptions that ensure that part of the book may be photocopied for research purposes without permission from the author.

Since their inception, not long after the development of the printing press, copyright laws have evolved to encompass literary, dramatic and musical works and new forms of transmission to the public such as broadcast and cable relay. [Fitzgerald 1999] provides an analysis of copyright law and concludes that the new digital environment presents serious challenges for existing intellectual property law.

According to [Fitzgerald 1999], information is a highly valued commodity in the new digital environment. However, fundamental features of information lead to difficulties in applying laws that regulate other commodities. As an intangible commodity, information can be easily copied many times. [Weber 1999] ironically notes the internet is the most efficient copying machine mankind has developed. Large amounts of information may be simultaneously copied in a matter of seconds over thousands of miles by many people. Furthermore, the internet, as a fundamental infrastructure device for the information society, is global in nature. This raises difficult jurisdictional questions for activities that frequently cross national boundaries.

The global nature of the internet and the ease with which works can be published and copied challenges existing copyright laws. In most jurisdictions, the author of a work is the creator of the work. However, compilations

such as multi-media works often bring together sound, images and text from a variety of sources. Traditional publishing houses of compiled hardcopy material ensure that each component is licensed for use. However, when publishing is as simple as uploading a file to the internet this requirement is left to individual creators. Furthermore, the file may be housed with an internet service provider that is based in a nation that is not a party to international copyright agreements and hence the illegality of unauthorised publication is unclear.

With hardcopy works, a public institution may easily house the work for any member of the public to peruse. Members of the public may browse the work and determine whether a copy is desired. In Australia, if the purpose is related to study then 10% of the work may be copied under the fair use exceptions of the Copyright Act. Although technically possible, copying more than 10% becomes extremely cumbersome and costly in real space. The same copying in cyberspace is trivial.

Commentators on the future of copyright law in cyber space disagree less on the nature of information or the internet than they do on the extent to which copyright law can remain appropriate in a digital environment. We can summarise arguments made concerning copyright law in cyberspace into three main camps.

(1) Authors in the first camp advocate an overhaul of existing copyright principles. This can be achieved by creating separate laws for the internet; a kind of Lex Internet where copying is actively encouraged and software is freely available.

(2) Those in the second camp claim that very little change is needed at all. Existing copyright principles are, by and large, adequate for the digital age though minor adjustments may be required.

(3) Authors in the third camp claim that copyright principles will need to be revised but not totally overhauled. Regulation of copyright will continue to be demanded but its implementation is more likely to be realised with the use of technology perhaps in conjunction with legislation or other measures.

The aim of the project described here is to advance a proposal for the regulation of copyright with the use of knowledge based systems. As such, we agree with authors in the third camp; that regulation of copyright will be largely realised with a combination of software, statutory and other measures. However, if the software is knowledge based and is designed to be integrated into a plausible business model, then concerns associated with technological regulation of copyright can, to some extent be alleviated.

In the next section of this paper we survey existing technological approaches that protect authors rights. Following that we examine arguments made by each camp above in order to conclude that a combination of technology and law is plausible particularly if software is designed to enable access to protected works for purposes of public benefit. We then describe five knowledge based systems that are sufficiently flexible to protect authors rights without denying the public access to works for fair use purposes. In the last section we describe the business model we see can be used to house the knowledge based systems.

2. Trusted systems

[Stefik 1997] defines a trusted system in the context of digital publishing as a system that follows rules governing access to a protected work. A trusted system determines whether a user has a right to access a work. If so, the system produces a copy for the user otherwise an error message is generated. Rights of different users are set by the author or owner of the work. The expression of rights can be very complex and include different categories of user, different parts of the work for a number of proposed uses at various time intervals for a range of payment options. According to [Stefik 1997], the expression of rights is so complex that a number of organisations including including Folio, IBM, NetRights, Xerox, and Wave Systems are developing formal languages for encoding digital rights. A trusted system must understand the rights information associated with a work in order to determine the conditions that are appropriate for the release of a work to any user.

A trusted system protects unauthorised access to a work with the use of two main types technological devices; locking and tagging. Detailed descriptions of these devices can be found in [McCulloch and Commins 1999], [Wayner 1997] and [Stefik 1997]. In locking systems, files are typically encrypted in such a way that only users the author specifically licenses have the necessary decryption keys. A common type of locking, called public key encryption involves the use of two keys, a public key and a private key. The public key is kept secret by the author of a work whereas a public key is freely available. A work encrypted with the private key can be decrypted with the public key and conversely, a work encrypted with the public key can be decrypted with the private key. The public/private keys are constructed in a manner that makes the derivation one from the other close to impossible.

Works locked with public key encryption are totally inaccessible by unauthorised users unless the key is cracked, for example, by attempting all possible keys, a feat that is enormously expensive computationally and close to impossible. A trusted system can encrypt a work with a secret private key and transmit the public key to an authorised user for decryption. However, under this kind of approach it is not possible for an unauthorised user to gain access to a digital work regardless of the intended purpose underpinning the access. The purpose of the intended use of the work becomes irrelevant so the unauthorised copying of a portion of the work for research purposes, for example, is impossible, even though this would be a legitimate fair use exception under copyright law.

The tagging approach involves incorporating a watermark or digital signature into a work so that the author is always recognised but the file is not blocked for use by unauthorised users. Watermarking devices applied to an image, for example, change a small number of pixels to insert an owners mark in a way that is imperceptible to a human observer. In the next section we survey arguments regarding the impact of cyber-space on copyright principles.

3. Copyright in cyber-space

As outline above, arguments made about copyright can be loosely grouped into three camps; the 'dispense with copyright' camp, the 'patch up existing law' approach and the 'integrate software with statutes' camp. In this section we outline arguments in each camp in order to claim that the latter point of view is more plausible. However before doing this we digress briefly in order to examine the nature of regulation in general.

3.1 Regulation

[Lessig 1999] identifies four types of mechanisms that regulate social behaviour; the law, market forces, social norms and natural phenomena. A simple example illustrates these mechanisms. Motor vehicle speed along suburban streets may be regulated using the law by the implementation of an ordinance that sets a maximum speed, appropriate signage to inform drivers, speed detection mechanisms and appropriate penalties for offenders. Motor vehicle speed may also be regulated with the installation of speed humps. In this case, no ordinance needs to be passed, no signage is needed and penalties are not relevant as the humps provide a natural barrier to speed. Similarly, the regulation of individual smoking could be realised by a legislative ban on smoking. However, experience with prohibition of alcohol suggests that the black market and other side effects of such an attempt are undesirable. Instead, smoking is more aptly regulated by market forces, by ensuring the price of cigarettes is kept extremely high in conjunction with extensive advertising campaigns that aim to alter social norms so that smoking becomes socially undesirable.

Natural barriers are proficient mechanism for regulation. As [Lessig 1999] points out there is no need for laws prohibiting the theft of skyscrapers because of the physical impossibility of stealing a building. In the physical world natural barriers such as humps on roads or unmoveable buildings are typically obvious. In cyberspace, natural barriers are implemented by software and are not so obvious. [Lessig 1999] uses an example about chat rooms organised by a large, global internet service provider. The number of participants in a chat room is regulated by software that admits users up to the maximum number and displays a message inviting others to try later. The regulation of participants in chat rooms using software restricting access is not as transparent as it would be if the regulation was implemented with laws, market incentives or social norms.

3.2 Dispense with copyright

[Stallman 1994] argues that copyright is inappropriate for current technologies. For instance, he advances the idea that the internet should remain a forum where creators relinquish rights existing copyright laws grant them and instead develop new business models in order to derive financial benefit from their creations. His arguments are compelling and attract additional support from his association with the Free Software Foundation (FSF), an organisation that is committed to the development of software for free public release. This foundation has developed numerous software tools such as the compiler, gnu c that are highly sought not only because they are free but because they can be freely modified to suit specific needs. The Free Software Foundation does not profit financially from the software it creates in the same way that a traditional company such as MicroSoft benefits from every sale of its copyright protected software. Instead, FSF adopts a business model that attracts profit from support services associated with its creations. Similarly, The Red Hat Corporation earns income by offering services that support the freely available operating system, Linux.

As appealing as arguments from this camp are, the same technology that led to the internet seems likely to diminish the strength of these arguments. Software such as locking devices outlined above, have already been developed that denies unauthorised use regardless of public

benefit. A widespread practice of allowing users to copy works freely is unlikely to emerge given the availability of locking devices. Authors faced with a choice of making their work freely available on the one hand, and on the other, encrypting it so that only authorised, and paying, customers gain access, are far more likely to opt for the encryption.

Legislation that attempts to prohibit the use of locking devices on the internet is difficult to imagine. This would require a separate law for the internet. As [Lessig 1999] notes there is as much chance of the emergence of a lex internet as there is for a law of the horse. Even if changes to existing copyright laws were immediate and global their enforcement would be impossible. The internet is an unregulated infrastructure so authors will find ways to employ blocking software.

3.3 Patch up existing law

Arguments in the second camp advocate minimal modifications to existing copyright laws. [Dixon and Self 1994], in an early attempt to anticipate cyber-space issues, argue there are five critical junctures of the information superhighway for purposes of copyright law; the creation of intellectual property, uploading content, transmitting content, accessing content and using content. They assess existing copyright laws at each juncture and conclude that existing laws are well-suited to protecting the interests of authors in the digital age. In stark contrast to authors in the first camp above, [Dixon and Self 1994] argue that continued vigilance with copyright laws is even more important in a digital age.

Undoubtedly, amendments to existing statutes need to be made to deal with radically new technologies but the changes do not alter, in any fundamental way, the structure of copyright protection. For example, the Copyright Amendment (Digital Agenda) Bill 1999 currently before Australian parliament replaces technology specific rights such as a right to wireless broadcast with a technology neutral right to transmit to the public. A fundamental position held by adherents to this position is that cyber-space may be a new communication medium but that does not alter principles underpinning copyright law. Any argument claiming otherwise is overstating the case.

Despite these claims warnings regarding a shift away from public benefit rights toward authors rights have been raised by [Clarke and Dempsey 1999], [Stefik 1994] and also by [Richter and Chicola 1999]. [Griffiths 1999] warns that the public interest aspects of copyright laws, typically implemented as fair use exceptions to authors rights are being lost in new amendments to European copyright laws.

3.4 Integrate software with statutes

We accept the view that effective regulation of copyright in cyber-space is likely to be realised with laws in conjunction with technological devices such as blocking or tagging software. We propose the development of technological devices that employ blocking software but do not deny access for public interest purposes. The software is necessarily knowledge based as it must apply knowledge about copyright laws to user supplied information regarding the intended use of protected works. In the next section we describe the knowledge based framework and then, in the following section, detail the representation of knowledge that we believe is central to the framework.

4. Copyright protection using a knowledge based framework

The five junctures [Dixon and Self 1994] identify as distinct points in cyber-space that concern copyright provides a basic framework for the approach presented here. These junctures; the creation of intellectual property, uploading content, transmitting content, accessing content and using content involve different agents and require different combination of technological, legal, economic and natural devices for regulation. We describe the actors in uploading content juncture and identify some of their needs in order to make the point subsequently that the needs are sufficiently diverse, complex and open textured to require a knowledge based system approach that is more sophisticated than rule based reasoning.

4.1 Uploading content

The second juncture that involves the application of copyright concerns steps taken to upload to a network such as the internet. The actors at this stage are the author/owners and internet service provider. Authors/owners need advice on:

- How to assign appropriate rights management information (RMI). RMI will ideally be expressed with the use a formal language developed for the purpose and mentioned previously.
- The best technological device for protection. Some works are best left unprotected because the author desires liberal distribution. Other works are suited to

watermark protection and others require public key encryption.

- The status of the internet service provider in relation to the work. For example, the system can advise the author of a text that contains numerous examples of bad language that the internet service provider will be required, under Australian legislation, to remove the work if a complaint is received.

Two questions emerge from the articulation of the needs of users at the two junctures above in relation to copyright law; can a knowledge based system be developed that meets those needs ?, and, will a knowledge based system be used even if it did meet those needs ? We shall offer an answer to the latter question first.

4.2 Will a knowledge based system be used ?

A knowledge based system will be used in the creation of the work phase if the system is trusted, and if the appropriate incentive is applied. We suggest that an appropriate level of trust can be engendered by developing systems at each juncture where the user is in control of the level of autonomy a system has as follows:

- Least autonomous system. Each least autonomous system is a stand-alone knowledge based system that offers advice to users. The system does not interact with other systems and performs no actions. The user is in total control of the software. For example, a least autonomous system for the Uploading content phase will help a user identify a package of rights that may be appropriate for a particular work but will not attach those rights to a work without prompting the user. Least autonomous systems must explain their reasoning in sophisticated ways. As [Buchanan et al 1995] note, the level of user acceptance of their medical expert systems depends heavily on the sophistication of explanations offered.
- Fully autonomous system. Each fully autonomous system interacts with other systems as much as possible. For example, the Uploading content system itself determines the appropriate level of rights, encodes those in a standard language and attaches them to the file. The same system itself determines the most appropriate encryption method and, interacting with encryption software, performs the action. A fully autonomous Using content system would prompt a user intent on accessing a work protected by the system for relevant information. The fully autonomous system would determine whether a copy ought to be released and if so, perform the necessary decryption.
- Semi-autonomous systems. Semi-autonomous system perform some functions autonomously and others under user guidance.

[Buchanan et al 1995] observe that the degree of control a user has over how software operates is an important aspect of user acceptance and use of their expert systems. Placing the user in control is a fundamental principle in the design of user interfaces according to Microsoft Corporation guidelines [Microsoft Corporation 1995]. Furthermore, as [Lessig 1999] notes, software that autonomously performs actions such as restricting the number of users in a chat room, introduces a device that regulates in a manner that lacks transparency.

Authors of works are unlikely to trust systems that are fully autonomous at the outset. However, authors who can control the degree of autonomy of each system can opt to hand over more and more actions to the system as they increase their understanding and trust of the system.

As articulated above, a knowledge based system will be used if the systems are trusted, and if the appropriate incentive is applied.

5. Argument based knowledge representation

The knowledge based systems that are proposed here need to draw on knowledge that encodes copyright legislation, precedent cases, international law and technological information. The systems need to be flexible so as to be easily maintained and extendable to changes in law or technology. The knowledge must be represented in such a way that the knowledge bases can be easily maintained. Furthermore, knowledge must be shared between the systems so as to reduce redundancy in storage and anomalies that can arise from updating parts of one knowledge base. The copyright advisory systems proposed here must explain their reasoning very well so that user's trust in their processes is ensured.

The knowledge representation scheme is central for any knowledge based system but is particularly important for this application. A structure based on arguments put forward by the philosopher [Toulmin 1958] is appropriate for our purposes. This structure has been used to represent arguments made in property disputes following divorce by [Stranieri et al 1999] and arguments made by applicants for refugee status in asylum law by [Yearwood and Stranieri 1999]. A variations to this structure has been used by [Bench-Capon 1998] in engaging users in a dialogue.

This Toulmin structure that [Stranieri et al 1999] facilitated the integration of various inference methods from artificial intelligence including neural networks, rule based reasoning or fuzzy logic. This is advantageous because as [Harris et al 1994] points out, hybrid architectures are often more flexible than single systems. The argument based representation facilitates the construction of explanations even when neural networks are used. Furthermore, [Stranieri and Yearwood 1999] demonstrate that the automated generation of documents that summarise the reasoning steps is facilitated if knowledge is represented using the structure. The representation is critical to our proposal and is based on philosophical insights that are discussed in [Stranieri *et al* 1999].

Despite the immediate appeal of TAS as a convenient frame for representing knowledge, most researchers that use Toulmin structures vary the original structure. A survey of variations can be found in [Stranieri and Zeleznikow 1999]. In our variation, we explicitly identify three features that are left implicit in the Toulmin formulation:

- an inference procedure, algorithm or method used to infer an assertion from datum
- reasons which explain why a data item is relevant for a claim
- reasons that explain why the inference method used is appropriate

Objectives of the creation support system mentioned above include advice regarding the extent to which a work is original. The system provides the advice by constructing the most plausible argument for the originality of the work. The system explains its reasoning by drawing upon the reasons for relevance and appropriateness warrant components of the Toulmin structure. At present this is done in a rigid way that is an example of a 'canned text' approach according to the explanation classification framework suggested by [Moore 1995]. However, preliminary research suggests that extending the sophistication of explanation beyond 'canned text' will be facilitated if the argument structure is used to represent knowledge.

The argument structure facilitates the decomposition of a problem into a series of inter-linked arguments. By decomposing a task into arguments that feed into each other enables the representation of quite complex chains of reasoning in a compact way. Family law property proceedings, considered too complex and discretionary to model at the outset of the Split Up project, were modelled using a total of 35 inter-connected arguments. Refugee law reasoning is notoriously complex yet this has been modelled in the same way using just over 200 arguments by [Yearwood and Stranieri 1999]. The use of a number of small arguments feeding in to each other facilitates the maintenance of the system.

Experience in family law has suggested that changes to reasoning chains are typically localised to a small number of arguments. For example, the introduction of domestic violence as a relevant consideration in property proceedings following a divorce is arguably, the most significant change in Australian family law during the last five years. This change impacted on only one Split Up argument by requiring that a new data item be added and a modification of the inference procedure to take into account the new item. We expect that technological or legal changes in the field of copyright law will similarly be localised to small numbers of arguments that can be modified relatively easily.

In refugee law, a decision maker is sometimes required to construct an argument that is not represented by one of the 200 modelled from past decisions. For example, an applicant for refugee status may claim that the 'LTTE harrass Tamils in Jafna' but offer no evidence or data items to support the claim. If this claim has not been previously encountered then the decision maker will seek out data items that may support (or rebut) the claim. To do so he/she will invoke search engines on sources as varied as Amnesty International reports and U.S Department of State reports. The structure facilitates the organisation of knowledge for the decision maker. Research is in progress to explore ways in which the precision/accuracy of an information retrieval engine can be improved with the use of context information from the Toulmin structure.

We expect that facilitation of the construction of a new argument will be a feature of the copyright knowledge based systems proposed here that is particularly important in the early years of copyright regulation. For example, new arguments that advance claims for activities that ought to be defined as 'research' in order to qualify for fair use exceptions are to be expected, particularly as 'fair use' is not defined by any of the main statutes.

6 Conclusion

The ease with which created works can be copied and transmitted to a global audience with new cyber-space technologies presents serious challenges for the regulation of copyright. A number of commentators argue that new economic models that include a weakening of copyright protections are needed while others argue that the

copyright law needs only minor modification in order to accommodate new technology. Others claim that the most appropriate way to regulate the use of created works is to combine existing copyright law with technological measures such as encryption software. Although not yet widespread, the use of encryption software in conjunction with the modification of copyright legislation seems to be emerging as the main mechanism for the regulation of copyright. However, the main disadvantage with this solution is that the free availability of works for public benefit purposes such as study, research, news or review is threatened as encryption software can all too easily lock all unauthorised access to works regardless of public benefit purposes.

We propose a framework that involves the development of five knowledge based systems that can protect authors works using a variety of technological measures such as encryption without restricting access to works for public benefit purposes. Though not as simple as exclusive locking systems, the knowledge based systems are likely to be used if users have sufficient trust in their performance and if appropriate incentives are made available.

The systems are designed in such a manner that the user can easily vary the extent to which actions such as encoding information regarded desirable rights for a work, are performed automatically. Authors can use systems in minimal autonomy mode until they have sufficient trust that systems are performing in accordance with expectations. When this occurs, authors can incrementally direct the system to be increasingly autonomous. The principle of keeping the user in control will engender sufficient trust in the systems. This, together with appropriate copyright legislation and economic incentives will encourage the use of these systems.

The central challenge in developing the systems is to identify a knowledge representation framework that will facilitate the generation of explanations and enable the modelling of complex copyright knowledge in a manner that is easy to maintain. We describe the application of a framework drawn from the structure of arguments proposed by the philosopher, Toulmin.

Acknowledgements

This research was funded by an Australian Research Council grant in collaboration with Victoria Legal Aid and Software Engineering Australia Ltd.

References

W Barambah, M., and Kukoyi, A., 1999. Protocols for the use of Indigenous Cultural Material. in Fitzgerald, A., Fitzgerald, B, Cifuentes, C and Cook, P.,(Eds) 1999. Going Digital. Legal Issues for e-commerce, software and the internet. Prospect Media. Australia. Pp131 - 137.

Bench-Capon, T. J. M., 1998. Specification and Implementation of Toulmin Dialogue Game. in Hage, J., C., Bench-Capon, T., J., M., Koers, A., de Vey Mestdagh, C.N.J and Grutters, C., (Eds). 1998. Legal Knowledge Based Systems. Jurix: The Eleventh Conference. Gerard Noodt Institut (GNI). The Netherlands.

Buchanan, B., Moore, J., Forsythe, D., Carenini, G., Ohlsson, S. and Banks, G., 1995. An Intelligent interactive system for delivering individualised information to patients. Artificial Intelligence in Medicine. 7 (1995) 117-154

Clark, P . 1991. A Model of Argumentation and Its Application in a Cooperative Expert System. PhD thesis. Turing Institute. Department of Computer Science. University of Strathclyde. Glasgow.

Clarke, R., and Dempsey, G. 1999. Electronic Trading in Copyright Objects and Its Implications for Universities. Australian EDUCAUSE'99 Conference, Sydney, 18-21 April 1999. At http://www.anu.edu.au/people/Roger.Clarke/EC/ETCU.html

Dixon, A., and Self, L., 1994. Copyright protection for the Information Superhighway. In Firth, A., Lane, S and Smythe, Y. (Eds) 1998. Readings in Intellectual Property. A selection of Articles from EIPR and Ent. L. R. Sweet and Maxwell. London.

Fitzgerald, A., and Cifuentes, C., 1999. Pegging out the Boundaries of Computer Software Copyright (Eds) 1999. Going Digital. Legal Issues for e-commerce, software and the internet. Prospect Media. Australia. Pp37 - 70.

Fitzgerald, B., 1999. Conceptualizing the Digital Environment in Fitzgerald, A., Fitzgerald, B, Cifuentes, C and Cook, P.,(Eds) 1999. Going Digital. Legal Issues for e-commerce, software and the internet. Prospect Media. Australia. Ppl - 21.

Griffiths, J. 1999. Holding back the Tide - a Review of Recent Developments in Copyright Law in the United Kingdom. International Review of Law, Computers and Technology. Vol 13, No 3. Pp 283-301

Harris, T., MacIntyre, J., Smith, P., and Medskar, L., 1994. Neural Networks and Expert Systems: Com-

plimentary Technologies that Can Work Together. in Milne, R., and Montgomery, A., (eds) Applications and Innovations in Expert Systems II . Proceedings of Expert Systems 94. Cambridge. SGES Publications. pp 249-259.

Lessig, L., 1999. Code and Other Laws of Cyberspace. Basic Books

Matthijssen, L. J., 1999. Interfacing between lawyers and computers. An architecture for knowledge based interfaces to legal databases. Kluwer Law International. The Netherlands.

McCullagh, A., and Commins, I., 1999. Cryptography: from information to intelligent garbage with ease. in Fitzgerald, A., Fitzgerald, B, Cifuentes, C and Cook, P.,(Eds) 1999. Going Digital. Legal Issues for e-commerce, software and the internet. Prospect Media. Australia. Pp205 - 214.

Microsoft Corporation. 1995. The Windows Interface Guidelines for Software Design. Microsfot Press.

Moore, J., 1995. Participating in explanatory dialogues : interpreting and responding to questions in context. MIT Press. Cambridge, Mass.

Perelman, C. and Olbrechts-Tyteca, L. 1969. The New Rhetoric. translated by Wilkenson, J. and Weaver, P. University of Notre Dame press. Notre Dame, Indiana. originally published in 1958 in French as Perelman, C and Olbrechts-Tycteca, L. 1958. La Nouvelle Rhétorique: Traité de l'Argumentation. Presses Universitaires de France.

Richter, K., K., Chicola, J, M., 1999. Digital Rights for Intellectual Property Protection. Proceedings of the IASTED International Conference. Law and Technology (LawTech'99). ACTA Press. pp27-31.

Stallman R. 1994 'Why Software Should Not Have Owners', at http://www.gnu.org/philosophy/why-free.html

Stefik M. 1997 'Shifting The Possible: How Trusted Systems And Digital Property Rights Challenge Us To Rethink Digital Publishing', Berkeley Technology Law Journal, 12, 1 (Spring 1997), at http://www.law.berkeley.edu/journals/btlj/articles/12-1/stefik.html

Stranieri, A., Zeleznikow, J., Gawler, M., and Lewis, B., 1999. A Hybrid rule- neural approach for the automation of legal reasoning in the discretionary domain of family law in Australia. Artificial Intelligence and Law Vol 7. Pp153-183

Stranieri, A., and Zeleznikow, J. 1999. A survey of argumentation structures for intelligent decision support. in Burstein, R. (Ed) Proceedings of International Society for Decision Support Systems Fifth International Conference ISDSS'99. Monash University, Melbourne.

Toulmin, S. 1958. The Uses of Arguments. Cambridge University Press. Cambridge

van Dijk, T. A. 1989. Relevance in logic and grammar (in) Norman, J and Sylvan, R (eds). Directions in Relevant Logic, pp25-57.

Wayner, P. 1997. Digital Copyright Protection. AP Professional. Academic Press. Chestnut Hill. MA

Webber, D, B., 1999. Intellectual property in internet. Fitzgerald, A., Fitzgerald, B, Cifuentes, C and Cook, P.,(Eds) 1999. Going Digital. Legal Issues for e-commerce, software and the internet. Prospect Media. Australia. Pp21 - 36.

Yearwood, J., and Stranieri, A., 1999. The integration of retrieval, reasoning and drafting for refugee law: a third generation legal knowledge based system Seventh International Conference on Artificial Intelligence and Law. ICAIL'99 ACM Press. Pp 117-137.

Case-Based Sentencing Using a Tree of Legal Concepts

Yaakov HaCohen-Kerner* ; Uri J. Schild**

*Department of Computer Science, Jerusalem College of Technology (Machon Lev);
21 Havaad Haleumi St., P.O.B. 16031, 91160 Jerusalem, Israel

**Department of Mathematics and Computer Science, Bar-Ilan University; 52900 Ramat-Gan, Israel
kerner@avoda.jct.ac.il ; schild@cs.biu.ac.il

Abstract

The Judge's Apprentice is a case-based decision support system implemented and intended for use in Israeli criminal law to aid sentencing in cases of either robbery or rape. The system uses a sentencing tree, which is an hierarchical classification of 371 legal concepts relevant to criminal sentencing. Each leaf in this tree represents an index, which can be input for any case. The indexes are important for determining index similarity in the retrieval process and for computation of the proposed sentence. After retrieval of suitable cases and selection of the best case, we use a case-based quantitative evaluation as a formula for computing a verdict for the case at hand.

1 Introduction

Case-Based Reasoning (CBR) is a powerful problem solving technique which enables solving new cases by utilizing the experience gathered in precedent cases. CBR is relevant for any intelligent domain where experience has an importance. It has been successfully employed in a variety of domains such as: mediation (Kolodner et al., 1985), law (Ashley, 1991), medicine (Koton, 1988), cooking (Hammond, 1986), navigation of an autonomous robot (Ram and Santamaria, 1997) and chess (Kerner, 1995A, 1995B). An extensive overview of CBR is available in Kolodner (1993).

The legal domain in general and the Anglo-American legal system adopted (with no jury) in Israel in particular, is an ideal domain for CBR since it relies on extensive utilization of previous precedents.

In a previous paper (HaCohen-Kerner and Schild, 1999) we presented a case-based sentencing model and the construction of its application in a computerized working system called The Judge's Apprentice. This system was developed for use in Israeli criminal law to aid sentencing in cases of either robbery or rape.

The focus of this paper is presenting the system's sentencing tree and the construction of the case-based computation for a proposed verdict for the case at hand.

This paper is organized as follows: Section 2 gives background concerning relevant sentencing systems. Sections 3 introduces the collection of indexes. Section 4 details the system's sentencing tree. Section 5 describes the case-based sentencing done in the model. Section 6 illustrates an illustrative example. Section 7 presents the results of the experiment and analyzes them. Section 8 summarizes the research and proposes future directions.

2 Previous sentencing systems

ASSYST (Simmon, 89) and LIST (Hogarth, 89) are legal systems that assist in actual uniform sentencing. LIST enables the judge to get various relevant information, e.g., statistics, legal files, and articles of law. ASSYST is a more radical system. It binds the judge to give a sentence in between an interval computed by the system according to predefined sentencing guidelines. However, these systems are not case-based.

JUDGE (Bain, 1989A, 1989B) is a CBR model that attempts to characterize how a judge decides on the sentence. It uses limited knowledge of a very specific type of crime and decides on the sentence with the aid of a simple formula without presenting explanatory details. JUDGE is a simple case-based model not intended for practical use.

3 Collection of indexes

Correct choice of indexes is critical for the retrieval of the most relevant precedents for the case at hand and for the computation of its sentence.

The indexes the judge use were found in the two main methods of formulating and collecting indexes presented by Kolodner (1993):

1. The functional approach: In this approach, indexes and data are gathered by examination of the cases at hand and the tasks needed to be carried out on them.
2. The reminding approach: indexes and data are gathered in this approach by analyzing and characterizing the various remarks of experts who carried out the proposed tasks in the past.

Most of the indexes, we collect were included in the model. They were divided into two main groups: facts (e.g., age, sex) and judge's evaluations (e.g., negative characteristics, offence with a potential of imitation).

However, there were two groups of indexes that have not been used since they were not practical for retrieval of cases and computation of the sentence. These groups were: informative indexes (e.g., type of court, name of judge) and indexes relevant for verdict-decision and not for sentencing (e.g., credibility of the offender and obsolescence of the offence).

All indexes are optional when inputting a new case. However, These indexes are stored as leaves in the sentencing tree described in the next section.

4 The sentencing tree

The sentencing tree is an hierarchical classification of 371 legal concepts relevant to criminal sentencing at different levels of abstraction. At the root of the tree, we have the concept of "criminal sentencing". Then we have four son-nodes: offence, offender, victim, and general. Each leaf (a node at the last level) in this tree represents an index (e.g., "knife", "boss", and "young").

This tree has been constructed by a research team with the help of a senior judge. Firstly, it has been constructed in the Ehrlich (1994). The tree has been improved and enlarged over a few years of this research. Indexes have been formulated and collected by the functional and the reminding approaches.

Figure 1 presents the three highest levels of this tree. Figures 2 to 6 illustrate most of its subtrees.

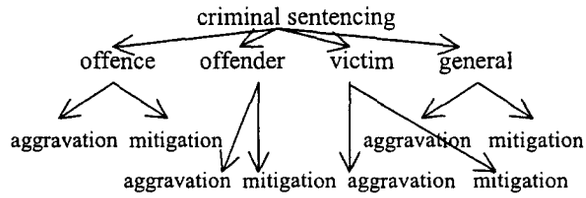

Figure 1: The three highest levels of the sentencing tree

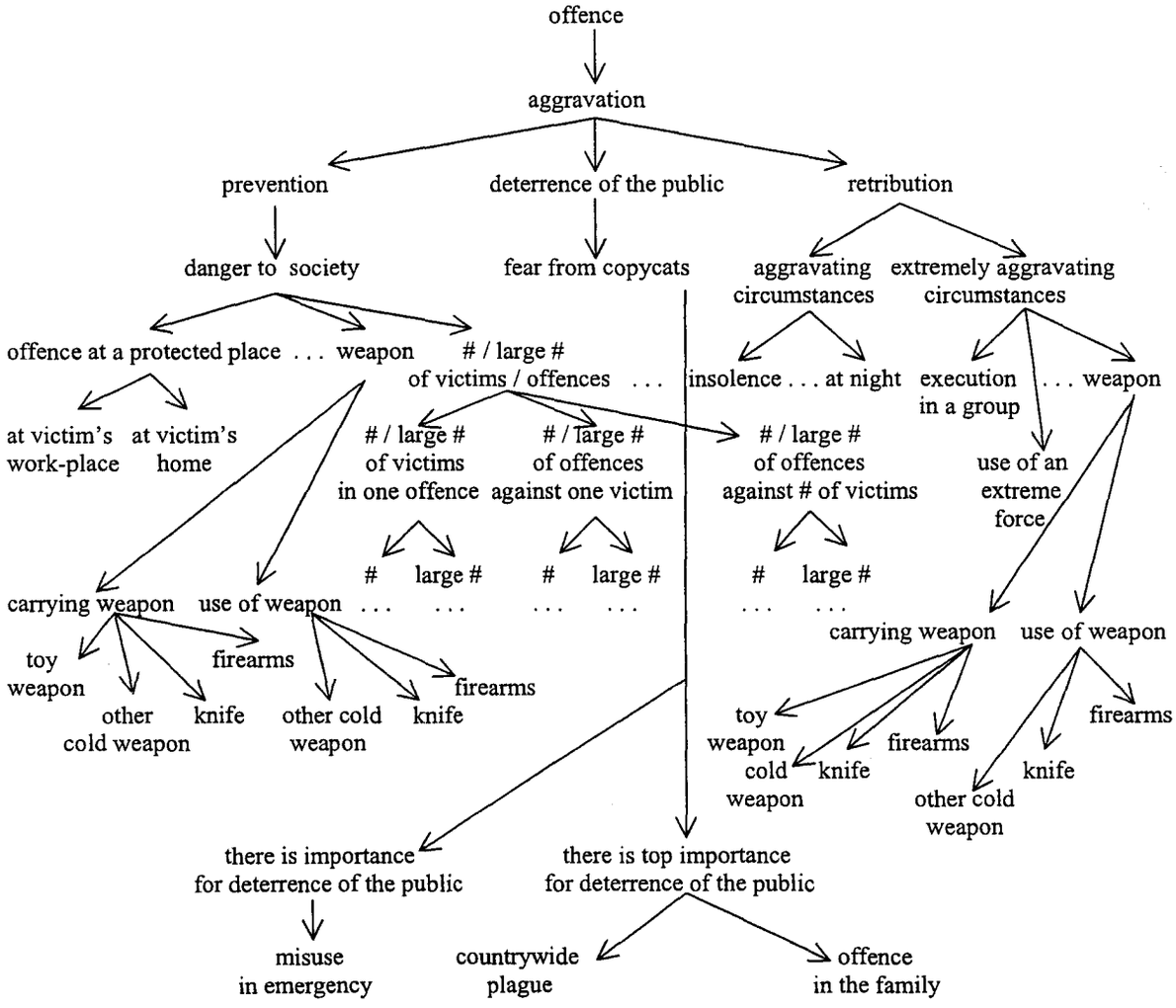

Figure 2: The subtree for the path "offence-aggravation"

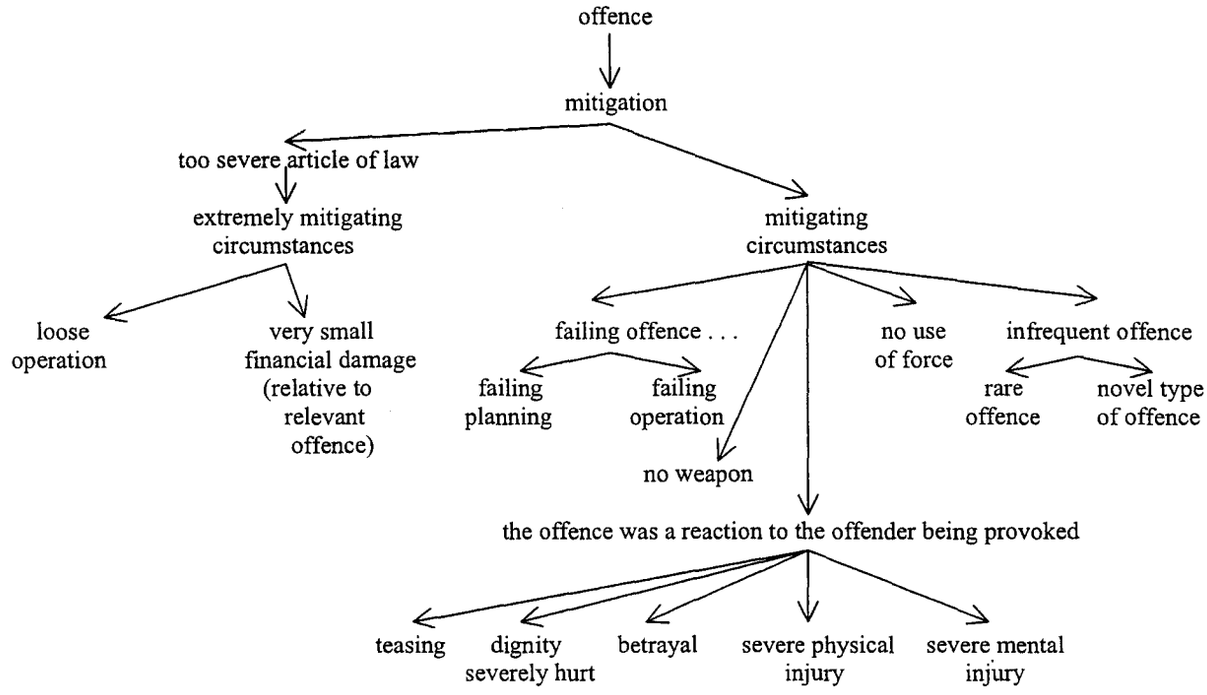

Figure 3: The subtree for the path "offence-mitigation"

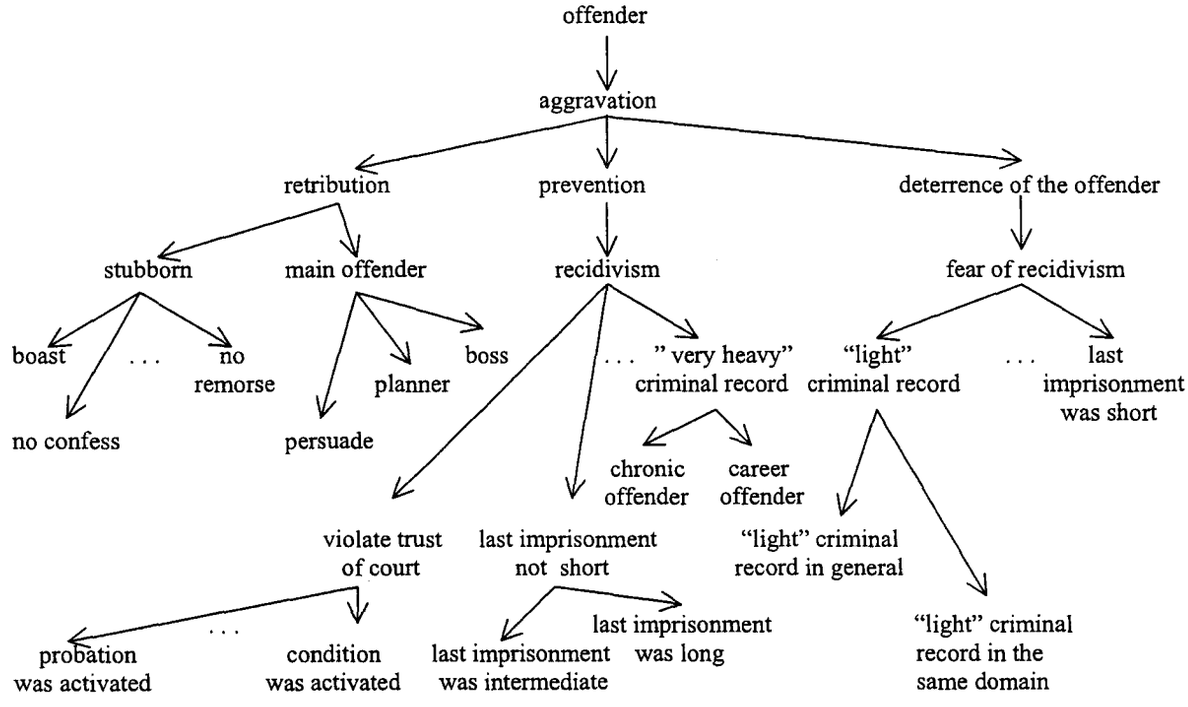

Figure 4: The subtree for the path "offender-aggravation"

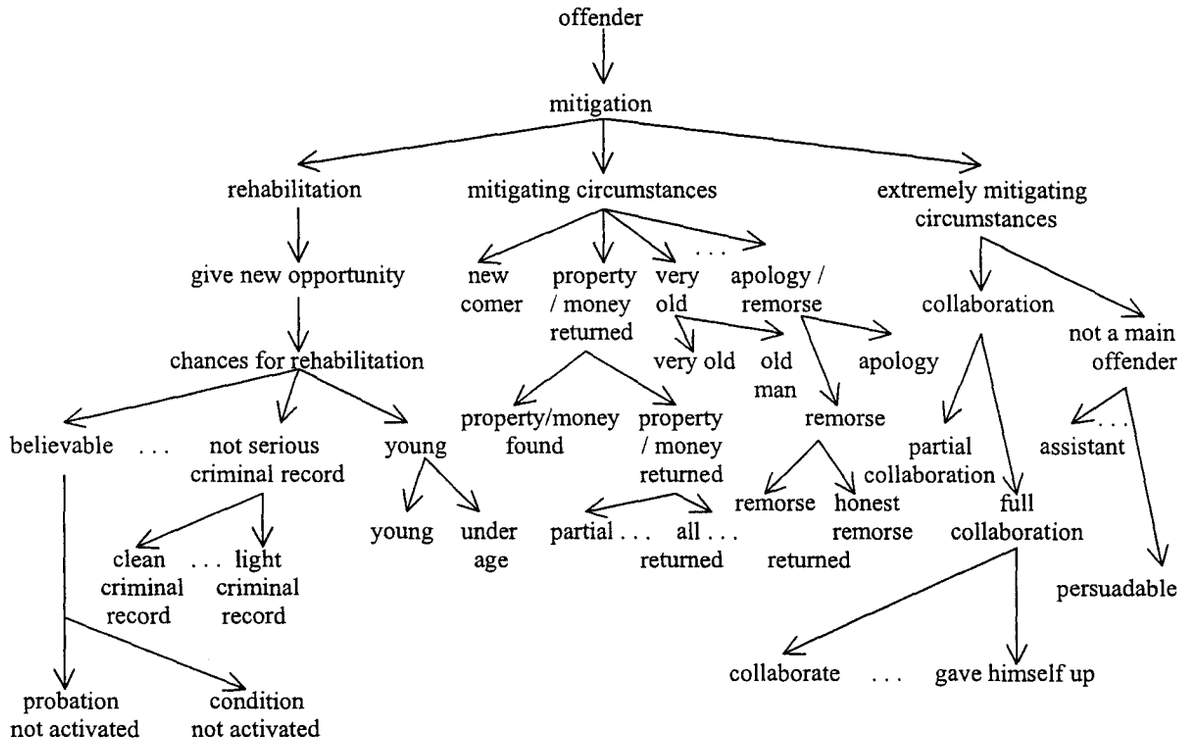

Figure 5: The subtree for the path "offender-mitigation"

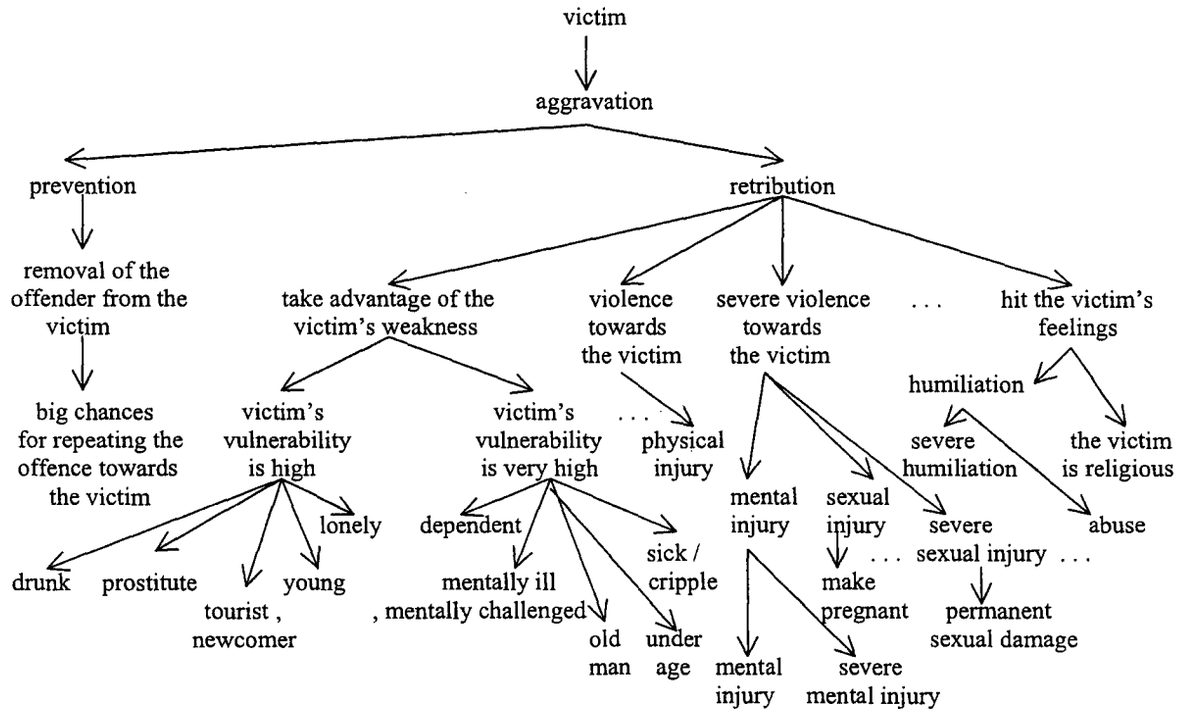

Figure 6: The subtree for the path "victim-aggravation"

5 Case-Based Sentencing

5.1 The algorithm

The case-based sentencing algorithm in our model is described in a detail in HaCohen-Kerner and Schild (1999). A summary of this algorithm is given below :

1. Inputting a New Case (NC) and its processing.
2. Initial retrieval of relevant precedents.
3. Retrieval of the most relevant precedents by a similarity measure.
5. If the retrieval was successful, the closest, most similar precedent is set as the Old Case (OC).
6. Constructing a solution for the NC by using the OC including a computation of sentence for the NC.
7. Limited learning has been enabled by acquiring the NC as a precedent at the judge's initiative.

The system's knowledge includes: a data base with fifty-four cases, legal knowledge relevant for sentencing (e.g., articles of law and maximum punishments), and a sentencing tree which is an hierarchical classification of 371 legal concepts relevant to criminal sentencing.

In the next subsection we shall describe in detail, the computation of the sentence for the new case based on the retrieved precedent is set as the Old Case (OC).

5.2 Computation of the sentence

5.2.1 Assumptions and definitions

Computing the sentence is the main step in our algorithm. We formulate three different kinds of formulas, which are presented in the next subsections.

In some of the formulas, we use the concept of the *base-sentence.* This is the sentence that would have been passed on the case without indexes towards severity or leniency. Computing the sentence for any case is based on an evaluation of the measure of difference, either more severe or more lenient, from the *base-sentence.* The *base-sentence* is presented by a positive whole number that describes the number of months of active imprisonment that would have been passed on a case of the above types.

In accordance with the above definition, for each type of criminal offense (robbery, rape) a separate *base-sentence* is defined. The values of the *base-sentence* (in months of active imprisonment) for each of the offenses that we deal are 36 months for robbery and 48 months for rape offenses.

In addition, we assume in all formulas the use the weight of the n most important indexes of comparison between the NC and the OC.

In the application, the value of n has been set at 7. Any number smaller than 7 would be too weak for comparison between the NC and the retrieved OC. Furthermore, the number 7 is accepted as the number of items that the average person is able to remember without apparent effort, according to the cognitive rule called "7±2". This means that the average person is capable of remembering approximately 7 information items (between 5 and 9) over a relatively short term (Miller, 1956).

W_i $(i = 1, \ldots, n)$ are the weights of these n indexes. A weight (also called general weight) of an index is defined by an integer number (either positive or negative) describing the influence of the discussed index on the *base-sentence.* This weight describes the change in percent in the *base-sentence,* assuming that this is the unique index that influences on the sentence.

Each leaf in the evaluation tree (Figures 1 to 6) has its own weight given by us as an arbitrary reasonable value. For instance the following leaves: "carrying warm weapon", "offence at victim's home", "offender is underage", and "confess" have the following weights +50%, +25%, -40%, -31%, respectively.

In the computation-formulas, we use some abbreviations as shown in Table 1.

Table 1: Abbreviations used in computation-formulas

abbreviation	concept
S_{NC}	the sentence computed for the NC
B	the base_sentence
S_{OC}	the sentence given to the OC

5.2.2 First version for the sentence

The first version for the NC's sentence is:

$$S_{NC} = B^* [\, 1 + \sum_{i \in NC} Wi(NC) / 100 \,]$$

This formula was rejected because of two reasons: (1) Often, using this formula, lead to extereme sentences according to those given in courts. The reason is that we do not take into consideration the interaction that exists betwwen legal indexes. (2) This formula does not use the retrived OC.

5.2.3 Second version for the sentence

The second version of the sentence computed for the NC uses the OC but not the *base-sentence* as follows:

$$S_{NC} = B^* [\, \sum_{i \in NC} Wi(NC) / \sum_{i \in OC} Wi(OC) \,]$$

Since, the OC is assumed as similar enough to the NC we assume that the NC's sentence can be computed by using the OC's sentence multipled by the propertion between the NC and the OC. This propertion is computed by the sum of weights of the n most impoertant indexes in the NC divided into the sum of weights of the n most impoertant indexes in the OC.

In addition, we take into account to some extent the interaction between the indexes of the NC by using the sentence given to the OC. We assume that the judge who gave this sentence took into account the interaction between the indexes of the OC.

However, this formula was rejected since the retrieved OC might be given before a long time and its sentence may be not relevant. To avoid some of its inflence we think we have to use the base-sentence. In contrast to the old case and its sentence, the base-sentence, which is updated by the judges is more reliable.

5.2.4 Third version for the sentence

The third version for the NC's sentence is:

$$S_{NC} = B * [\ 1 + comparison\ (NC, OC)\ ]$$

The current computation of the sentence for the NC is based not only on the base-sentence but also on the comparison between the NC and the retrieved OC. In this comparison, we take into consideration the weight of the n most important indexes found in the NC and the OC. The weights of these n indexes are computed with the aid of a complex process, which is described in detail in Hacohen-Kerner's (1997) doctoral project.

In general, this comparison is composed of four different values: (1) weight of identical important details found in both cases, (2) weight of important details found in the NC that have similar details to them in the OC, (3) weight of important details found in the OC that have similar details to them in the NC, (4) weight of the details found in the NC but are neither in the OC nor have similar details in the OC.

6 A Short Example

In this section, we illustrate the use of the Judge's Apprentice. The chosen NC is a robbery case (file-number: 2/11) whose abstract is as follows: A youth helped an elderly lady to her home. There, he beat her head and body with harsh blows using a chair. The elderly lady suffered from physical injury (a broken arm and wounds) and psychological injury. The youth stole 350 NIS (current currency) from her purse and fled. The accused had an extensive history in the same field.

The summary of the first precedent (file number: 919/80) retrieved for the NC from the precedents that have at least one of the principles of punishment policies chosen for the NC (punishment, deterrence, rehabilitation) is as follows:

A thief was found guilty of gang robbery of an old woman of about 80 in her home. The robbers, carrying a knife, with nylon socks over their heads, broke into the house, beat the old woman and stole the sum of 13,000 LI (older currency than the NIS). The old woman was seriously injured both physically and psychologically. The accused had a chronic history of crime. Two suspended sentences of an inclusive period of 22 months were enacted against him. The judges decided that the crime was repulsive and that it was a blow of nationwide significance that could be copied by others. The accused was sentenced to 84 months of active imprisonment + the 22 months of suspended sentence that were enacted, a total of 106 months of active imprisonment.

Table 2 Presents the seven most important points of comparison between the NC and the retrieved OC.

Using the third version for the NC's sentence, the system proposes for the NC a sentence of 103 months of active imprisonment, while the sentence that was passed for the OC was 106 months. The percentage of relative change is 3%. That is to say, the proposed sentence is very slightly more lenient in comparison to the sentence for the OC.

7 Research Results

A sample run of the system was demonstrated before four senior legal experts who were very positively impressed by it. They said that the system was intelligent and very interesting and could definitely be of advantageous use to judges. Three of them were of the opinion that the idea of using CBR for sentencing is very relevant and should be applied.

The results of the model's application will be able to undergo authentic examination and evaluation only when the knowledge-base is constructed by an accepted legal body. Since this knowledge is not yet available, the system was tested on a non-typical store comprising fifty-four authentic precedents. A test was carried out for each precedent comparing the sentence suggested by the system and the sentence that was actually passed.

Table 2: Comparison of Details between the NC (2/98) and the OC (919/80)

Main Identical details	
	Detail
1.	Repulsive act
2.	Victim is elderly
3.	Psychological injury

Main Similar Details			
NC Detail	**OC Detail**	**Source of Detail**	**Comparison**
Extensive history in same field	History of major offences	Heavy criminal history	More Lenient
Use of cold weapon	Carrying a knife	Weapon	More Serious
Use of excessive force	Use of force	Use of force	More Serious

Main Details in NC but not in OC		
	NC Detail	**Comparison**
1.	Dangerous injury	More Serious

Main details in OC but not in NC	
OC Detail	**Comparison**

Table 3: Results of similarity measure between the sentences according to the similarity measure in the retrieval

Type of retrieval	**First precedent retrieved of high similarity (At least)**			**First precedent retrieved of medium similarity**			**No precedent retrieved**
Success/ Failure	Success with first precedent	Success after recovery	Failure after recovery	Success with first precedent	Success after recovery	Failure after recovery	Failure
No. of cases	12	4	5	6	3	10	14
Total	19			21			14
Percentage	57%	19%	24%	31%	16%	53%	100%

A sentence for the NC (constructed by the system in accordance with the retrieved precedent) is compared to the sentence that was actually passed for the NC (since this is an authentic case). Success is considered when the difference between the percentage of the sentence for the NC, which was constructed by the system, from the actual sentence passed, is small. In the current version, the difference is defined by the interval (-0.33, 0.33). Failure is defined by when the difference is greater in either direction.

The results of similarity measure between the sentences according to the similarity measure in the retrieval are shown in Table 3.

A number of conclusions were gleaned from these results:

1. The store of precedents available does not cover the domain of cases that it is supposed to encompass. For 14 cases (over one quarter of the cases) no relevant precedent was retrieved. Moreover, for 15 other cases no relevant case was retrieved in practice, because the cases that had been retrieved for them were not relevant having led to an irrelevant result.
2. Attempts at recovering did improve the results (3 + 4 = 7 cases). However, this is an improvement of only 13%. It must be noted, however, that A. not all of the possibilities at recovering were attempted, that is we did not try to develop a solution by relying on the third retrieved precedent onwards nor on appeals of the second precedent onwards. B. As we have said, the store of precedents is not sufficiently encompassing.
3. As a result of the limited attempts at recovering that were carried out, it became clear that for almost one half of the cases (25 out of 54, a little over 46%), solutions were constructed that were considered successful. This is a reasonable result considering running a prototype based on a non-representative store of precedents.

The system reveals far better results for cases for which a first precedent of a relatively high similarity is retrieved. This means that a case can be solved with greater success if the precedents retrieved for it are of a relatively high similarity. The results were as follows: The system revealed a relatively high success rate (approximately 76%=57%+19%) including recovery, for cases for which the first precedent retrieved was of at least of high similarity. The system revealed a lower success rate (approximately 47%=16%+31%) including recovery, for cases in which a first precedent of medium similarity was retrieved.

8 Summary and Future Work

We apply a case-based decision support system for use in Israeli criminal law to aid sentencing. This system

uses for aims of retrieval and computation a hierarchical tree that includes 371 legal concepts relevant to criminal sentencing.

A number of proposals for future research are suggested. One is an elaboration of the model for the handling of other offenses, such as, multiple offenses, other types of offences, or the presentation of statistical information. Another is an elaboration of the model to other domains requiring case-based evaluation or to learning of different components in explanation patterns, by using creative adaptation strategies in domains less conservative than law.

Similarly, a cognitive study of the method by which judges make their decisions during the retrieval and reasoning processes could be made. Another idea is the application of an automatic learning process concerning tuning coefficients of the retrieval and sentence formulas, selecting cases, recovering from failures, etc.

Notes

The names Y. HaCohen-Kerner and Y. Kerner refer to the same person. The research presented here is mainly based on Hacohen-Kerner's doctoral project, which was supervised by Dr. Uri J. Schild and Prof. Martin Golombic at Bar-Ilan University. The dissertation was approved by the Senate of Bar-Ilan University on 1998.

Acknowledgments

Thanks to the judges who help us with relevant sentencing background, especially to Moshe Talgam, vice-president of the district court in Tel-Aviv. Thanks also to Ephraim Nissan for valuable comments.

References

K. D. Ashley. Reasoning with Cases and Hypotheticals in HYPO, *International Journal of Man-Machine Studies*, Special Issue: *AI and Legal Reasoning*, part 1, E. Rissland, (Ed.), 34:753-796, 1991

W. M. Bain. JUDGE. In: Riesbeck C. K. and Schank R. C., (Eds.), *Inside Case-Based Reasoning*, Lawrence Erlbaum Associates, Hillsdale, NJ, 93-140, 1989A

W. M. Bain. Micro JUDGE. In: Riesbeck C. K. and Schank R. C., (Eds.), *Inside Case-Based Reasoning*, Lawrence Erlbaum Associates, Hillsdale, NJ, 141-163, 1989B

G. Ehrlich. *Decision Support System in Criminal Sentencing*. Master Thesis (In Hebrew), Department of Mathematics and Computer Science, Bar-Ilan University, Ramat-Gan, Israel, 1994

Y. HaCohen-Kerner. *The Judge's Apprentice: A Case-Based Decision Support System for Criminal Sentencing*. Ph.D. Dissertation (In Hebrew), Dept. of Mathematics and Computer Science, Bar-Ilan University, Ramat-Gan, Israel, 1997

Y. Hacohen-Kerner and U. J. Schild. The Judge's Apprentice. *The New Review of Applied Expert Systems*, 5:191-202, 1999

K. Hammond. CHEF: A Model of Case-based Planning. In *Proceedings of the Fifth National Conference on Artificial Intelligence*, 267-271, Menlo Park: AAAI Press/The MIT Press, 1986

J. Hogarth. *Sentencing Database System*, User's Guide, List Foundation, Vancouver, British Columbia, 1989

Y. Kerner. Case-Based Evaluation in Computer Chess. In Haton J. P.; Keane, M.; and Manago M. (Eds.), *Proceedings of the Second European Workshop, EWCBR-94*, Lecture Notes in Artificial Intelligence 984, 240-254, Berlin: Springer-Verlag, 1995A

Y. Kerner. Learning Strategies for Explanation Patterns: Basic Game Patterns with Application to Chess. In Veloso M.; and Aamodt, A. (Eds.), *Proceedings of the First International Conference, ICCBR-95*, Lecture Notes in Artificial Intelligence 1010, 491-500, Berlin: Springer-Verlag, 1995B

J. L. Kolodner, Simpson R. L., and Sycara-Cyranski K., A Computer Model of Case-Based Reasoning in Problem Solving, *Proceedings of the Ninth Int. Joint Conference on Artificial Intelligence, IJCAI-85*, 284-290, Menlo Park, Calif., 1985

J. L. Kolodner. *Case-Based Reasoning*, Morgan Kaufmann., Inc., San Mateo, Calif., 1993

P. Koton. Reasoning about Evidence in Causal Explanations. In *Proceedings of a Workshop on CBR*, 260-270, San Mateo: Morgan Kaufmann., 1988

G. A. Miller. The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity of Information, *Psychological Science*, 63:81-97, 1956

A. Ram, and Santamaria J. C. Continuous Case-Based Reasoning, *Artificial Intelligence*, 90:25-77, 1997

E. Simon and G. Gaes. ASSYST - Computer Support for Guideline Sentencing, *2nd International Conference on AI and Law, ICAIL-89*, ACM Press, 195-200 ,1989

A. Tversky. Features of Similarity, *Psychological Review*, 84(4):327-352, 1977

Truth And Consequence: Complementing Logic with Values in Legal Reasoning

T.J.M. Bench-Capon
LIAL - Legal Informatics at Liverpool
Department of Computer Science
The University Of Liverpool
Liverpool

tbc@csc.liv.ac.uk

Abstract

In this paper I discuss attempts to model legal reasoning with case precedents. I draw attention to the need to consider not only the factors in a case, but also the social purposes that are served by deciding cases with reference to these factors. I show how considering such purposes allow us to give more predictive power to precedents; to discriminate between arguments advanced on the basis of factors; and to explain how the law can evolve as social attitudes change.

1 Introduction

In the introduction to Perelman and Olbrechts-Tytecha (1969), the authors write:

> "Logic underwent a brilliant development during the last century when, abandoning the old formulas, it set out to analyze the methods of proof effectively used by mathematicians. Modern formal logic became, in this way, the study of the methods of demonstration used in the mathematical sciences. One result of this development is to limit its domain, since everything ignored by mathmaticians is foreign to it. Logicians owe it to themselves to complete the theory of demonstration obtained in this way by a theory of argumentation" (p10).

This mathematically inspired model of reasoning has also dominated attempts in Artificial Intelligence to handle practical situations where we find that our theory contains conflicts, or our information is incomplete, so that we need to reason in a defeasible or non-monotonic manner. So far the ideas of Perelman have been picked up mainly by those interested in the natural language presentation of reasoning. In for example Grasso et al (2000) and Reed (1999) a theory of argumentation has been used to produce more convincing presentations of reasoning. The question arises, however, as to whether this is simply a matter of presentation, or whether the theory of argumentation impacts also on the conclusions that can be reached. In this paper I want to consider for one domain of reasoning, namely reasoning in law, whether we do indeed need to complete our theory of demonstration by a theory of argumentation.

2 Legal Reasoning

Legal reasoning takes place in a context provided by previous decisions. Although a framework is provided by statutes, the law is developed through a series of decisions applied to particular cases. Important is the doctrine of *stare decisis*, in which previous decisions are held to be binding on courts of equal or lower status. Thus the task confronted by a legal reasoner is to apply the law to a particular set of facts presented in the case under consideration in a way which is consistent with the way it has been applied to other sets of facts in the past. Noteworthy are the following points:

- The are only two possibilities for the decision: either the court must find for the plaintiff or for the defendant;
- One of the options must be embraced: a decision in favour of one party or the other must be made;

- The decision must be justified:the decision comprises not only the verdict, but a description of the case and the reasons for the decision;
- Disagreement is inherent in the system: the case only arises because there is disagreement, and the parties are represented by lawyers who will present the reasons why their clients should win. Moreover higher courts recognise that the outcome need not be unanimous. In the UK three judges sit in the Court of Appeal and five in the House of Lords (the highest court in the UK), and it is possible that the verdict is a majority one, with the right for dissenting judges to present their reasons for dissent;
- Disagreement may not be based on any difference in facts or the interpretation of fact. It may well be possible to agree on the issues pertaining to a case and yet disagree about the right outcome;
- The decision is defeasible: the verdict may be appealed to a higher court (except in the case of the highest court).

Perelman emphasises this possibility of disagreement (Perelman 1980):

> "if men oppose each other concerning a decision to be taken, it is not because some commit an error of logic or calculation. They discuss apropos the applicable rule, the ends to be considered, the meaning to be given to values" (p150).

The next section will discuss how legal reasoning has been modelled, and the extent to which logic can represent the reasoning.

3 Modelling Legal Reasoning

Attempts have been to model legal reasoning with cases using rules. The idea is to summarise the various decisions in the form of rules, and then to apply these rules to the new case in hand. This approach has been criticised both from a jurisprudence theoretical standpoint (e.g. Moles (1987)) and from a practical standpoint. It enshrines a positivist approach to law which makes the application of law look more mechanical that is felt appropriate and imposes a particular interpretation (which has no legal standing) on a decision. In practical terms the difficulties arise when attempting to discover the operative rule: there is the problem of the appropriate level of abstraction from the case facts to the terms to be used in the rule (e.g. Twining and Miers (1976)) - too little abstraction and the rule fails to apply to new cases, too much and inappropriate cases are brought under its scope - and there is also the problem that plausible rules provide too little coverage to be useful in predicting the outcome in new cases. A more popular approach has therefore been to employ case based reasoning.

The simplest approach to case based reasoning (e.g. Kowalski (1989)) is to identify a set of factors that may be present or absent in a case and describe the cases in terms of these factors. A new case is then matched with the existing cases, and the decision in the closest match is applied to the new case. This approach is somewhat unsophisticated, and does not seem to reflect the way cases are actually used in reasoning. The best, and most well known, attempt to model legal reasoning with cases is the HYPO system of Rissland and Ashley, best described in Ashley (1990).

In HYPO the factors under which a case is described are not surface features of the case, but issues and distinctions that have been employed in past decisions, and which result from a thorough analysis of the law. HYPO takes as its domain US Trade Secrets Law and examples of such issues would be the number of people who knew the putative secret, and the ease in which the secret could have been discovered by reverse engineering. Each of the factors represent a distinction which was introduced in some particular case, and which has been refined in succeeding cases. All of the issues will favour either the plaintiff or the defendant. If the plaintiff took measures to protect his secret and restrict its knowledge that will be in his favour. If the secret is easily reverse engineerable, that will favour the defendant.

Given these factors, and a representation of past cases in terms of them, when confronted with a new case we can organise the existing cases according to the factors shared. If we are now constructing an argument for the plaintiff we will ideally cite a case in which:

1) the decision was for the plaintiff;
2) the pro-plaintiff factors in that case are in the new case;
3) the pro-defendant factors in the new case are in that case.

If we can find such a case, the decision is clear - the new case is completely on point. Typically, however, there will be no such case, and either (2) or (3) cannot be satisfied. Here the defendant can argue that the cases are distinguishable, on the grounds that the plaintiff's case is weaker than the original if (2) is not satisfied, or that the defendant's case is stronger in the new case if (3) is not satisfied. Alternatively the defendant will be able to produce a counter example - a case decided for the defendant which is as on-point than the case cited by the plaintiff.

The plaintiff can now respond to these objections by distinguishing the current case from any counter-examples, or by pointing to cases where favourable factors missing in the current case did not prevent a decision for the plaintiff, or by pointing to additional factors present in the new case which favour the plaintiff.

HYPO is a sophisticated system, which does seem to reflect and enable moves typically made by lawyers when arguing with cases (citing, distinguishing, supplying counter examples, giving instances where a missing factor is not fatal, etc). While, however, it can reproduce the arguments which can be constructed by the different sides, it provides no insight into why one of the arguments might be favoured over the other.

An interesting presentation of HYPO can be found in a reconstruction of the reasoning by Prakken and Sartor[1] (Prakken and Sartor (1998)), which expresses a set of decisions in terms of a set of arguments. They represent a case as a conjunction of three rules:

1) conjunction of pro-plaintiff factors in the case is the reason to decide for plaintiff
2) conjunction of pro-defendant factors in the case is the reason to decide for defendant
3) either the pro-plaintiff reason is preferred to the pro-defendant reason or vice versa, depending on how the case was decided.

They can then describe the various argument moves in terms relating to these rules.

If we adopt Prakken and Sartor's representation of cases, we can conveniently represent our current understanding of the case law as a directed graph, following Bench-Capon (1999). All possible combinations of pro-plaintiff factors form a straightforward hierarchy ordered on the assumption that an additional factor for a side will strengthen the case for that side. The possible combinations of pro-defendant factors form a similar hierarchy. The two hierarchies are then joined by edges between them. An edge is inserted to connect the node containing the pro-plaintiff factors in a particular case with the node containing the pro-defendant factors in that case, and the edge is directed to indicate the dec sion made in that case. An example, with pro-plaintiff factors A and B, and pro-defendant factors C D and E, is shown in figure 1.

[1] In fact their formalism is capable of more than this. Indeed they specifically recognise that most on pointness is not the only source of rule priorities.

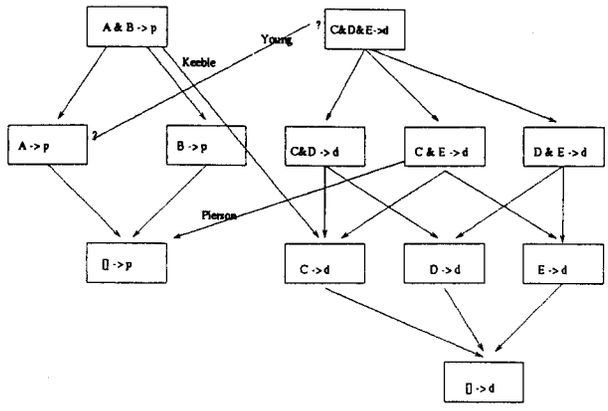

Figure 1: Graph Representation of Decided Cases in a Legal Domain: Pierson decided for the defendant, Keeble for the Plaintiff and Young under consideration

When we consider a new case we consider the two nodes representing the factors present in that case. If a path exists between them, the direction of the path will give us the decision determined by the precedents. The graph must contain no cycles, since this would imply that precedent determines victory for both sides - which would render our theory of the law incoherent. If no path exists we add an edge linking the two relevant nodes. Since the edge can be directed in either direction (no cycle can be introduced because no path currently exists) we can see that the outcome is not determined by the precedents currently represented by the graph. Eventually paths will exist between all pairs of nodes, when we can be said to have a complete understanding of this area of law. Such a graph may be said to represent a theory of law in this domain. It is *coherent* if acyclic, and *complete* if a path exists between every pair of nodes.

4 Limitations

The benefit of the modelling described in the previous section is that it can give a good *ex post* account of how a decision was reached: but it gives no account of *why* the decision was reached. Quoting Perelman (1980) again:

> "With the judicial syllogism we lose sight of the fact that the judge's intellectual effort has already been achieved, his deliberation finished, and there remains only the question of form. The important thing is not the passage from premises to conclusion but the way the judge justifies his premises, both in fact and in law" (p150)

If we know the facts of a case and the priority that is given to one rule as against another we can deduce the

result. But, no guidance is given as to why one rule should be preferred against another. Logic is, after all, concerned only with things known from their form alone: why then should be use it to predict things which depend on content? Thus the first criticism is that the account can explain past cases but when confronted with a new case, the decision appears arbitrary, whereas we know that it is held to be rational, and indeed reasons must be given by the judges. It is precisely this that Perelman requires his theory of argumentation for Perelman (1980):

> "A demonstration is correct or incorrect, it is imposed absolutely or lacks value; but in argumentation it is always possible to plead for or against, because arguments which support one thesis do not entirely exclude those supporting the other one; this in no way means that all arguments are of the same value. ... When we must justify preferences, deliberate about a decision, or discuss values, then argumentation and recourse to dialectics are indispensable." (p150).

The position both respects logic - it can rule out an argument as incoherent - but also recognises the need for rational criteria for the justification of two equally (logically) tenable views. They may be tenable logically while untenable with respect to dialectics.

Also, importantly, the taking together of all the factors in favour of a given side in a case fails to distinguish between several different situations.

1) It may be that two factors in fact require one another to have effect. Thus losing either of them would deprive the other of its weight in the argument;
2) It may be that two factors are entirely independent of one another. In such a case what is being presented when both are present is two distinct arguments, rather that one argument strengthened by the presence of both factors;
3) It may be that two factors can be substituted for one another; thus whether both are present, or either one, presents an equally strong argument.

The last point is addressed in Aleven (1997), where the notion of abstract factors is introduced. If two factors are subsumed under the same abstract factor, it is possible to argue that a case with one present has the same weight as one with the other present. This is an interesting step on the road to assessing arguments which are not obviously constrained by the theory: we can block, for example, a move which distinguishes the case on the absence of a given factor if another factor relating to the same abstract factor is present (irrespective of whether this factor was present in the original case).

The other two situations have, however, been no more than partially addressed: it may be that the first case above suggests the needs for sub-arguments, a possibility recognised without being discussed in detail, in Prakken and Sartor (1998). The second case, however, is not addressed at all in the literature: there is no way of distinguishing between factors making a single, stronger argument from those making two weaker ones.

The account we are looking for should accommodate all of these distinctions, and build on the notion in Aleven (1997) so as to extend our ability to assess logically coherent disagreements.

5 Values, Purposes and Consequences

The idea is that disputes which cannot be resolved by pointing to the logical coherence of the theory can be resolved by a consideration of the values of the parties to the dispute. Work such as that of Grasso et al (2000) has argued that to be persuasive an argument must conform not only to logic, but also to the values of those to whom it is addressed. In the case of a legal dispute judges attempt to provide a persuasive argument for the decision they wish to give, addressed in the first instance to the other judges trying the case with them, and then in the second instance to the public to which they are ultimately responsible. In a sense the judges are intended to embody society at large, and to try cases against the background of the value system of that society.

This is broadly the position taken by Moles (1987) who has criticised rule based representations of law as adopting an unrealistically formal approach. He argues that judges must consider consequences as well as the letter of the law.

The point is well made by an example quoted in Perelman (1980). Suppose the law has decided that dogs should not be allowed in a railway carriages. One should feel aggrieved neither if one is prevented from taking ones bear on board (even if it is on a leash), nor if one's neighbour is permitted access with his chihuahua: in both cases the purposes served by the ban on (typical) dogs are served also here.

One paper which has argued for the importance of representing the purposes underlying decisions is Berman and Hafner (1993). Their concern is to argue that case based models which aspire to reflect "the way in which practic-

ing professionals use legal decisions" must represent the purposes behind the rules articulated in cases. They centre their thinking on an example of the law as it develops through three cases: *Pierson, Keeble* and *Young*.

In the first case, *Pierson v Post*, the plaintiff was hunting a fox in the traditional manner using horse and hound when the defendant killed and carried off the fox. The plaintiff was held to have no right to the fox because he had gained no possession of it. In the second case, *Keeble v Hickeringill*, the plaintiff owned a pond and made his living by luring wild ducks there with decoys and shooting them to supply a local poulterer. Out of malice the defendant used guns to scare the ducks away from the pond. Here the plaintiff won. In a third case, *Young v Hitchens*, both parties were commercial fisherman. While theplaintiff was closinghis nets, the defendant sped into the gap, spread his own net and caught the fish. In this case the defendant won.

If we wish to argue *Young* against the background of *Pierson* and *Keeble* in the manner of HYPO (Ashley (1990)) we can readily identify a number of factors (which I associate with the letters shown in Figure 1). Whether the plaintiff owns the land (B), whether the land is open (E), whether the animal was caught (C) and whether either or both the parties are engaged in making their livelihood (A for the plaintiff and D for the defendant) are obvious ones. *Pierson* now looks rather clear: the plaintiff does not own the land, did not catch the beast and was looking for pleasure rather than business. No factor favours the plaintiff in this case. (Note that in part it is the choice of factors that lead to this result. The factors recognised in a domain reflect the decisions that have been made.) In *Keeble*, although the plaintiff was not in possession of the ducks, he had the pro plaintiff factors that the land was owned and he was engaged in a commercial pursuit, and we know that these were sufficient.

If the defendant were to propose that *Pierson* is followed in *Young* we can distinguish on the grounds that the plaintiff is making his living, and cite *Keeble* to show that not having captured the prey is not fatal to the plaintiff's case. But for the defendant we can distinguish *Keeble*, because in *Young* the water is not owned by the plaintiff, and add that the motive of the defendant was that he too had a living to make. So the issues are identified, but the question remains as to whether we are persuaded (or are able to persuade a judge) that the extra factors in *Young* are sufficient to cause us to reject *Pierson*. What makes one side of the case more persuasive than the other?

Figure 1 shows the situation in the notation described in section 3 above. If, however, we probe deeper into the decisions and find out why the factors used above are relevant, we can come to a clearer view of the matter. The judges who found in favour of the defendant in *Pierson* did so:

> "For the sake of certainty, and preserving peace and order in society. If first seeing, starting or pursuing such animals .. . should afford the basis of actions it would provide a fertile source of quarrels and actions." (Quoted in Berman and Hafner (1993)).

Although one judge dissented in *Pierson* (for him the pursuit and destruction of foxes was sufficiently socially valuable to be encouraged and protected by law), the majority line was clear: this was not the sort of dispute in which the law should concern itself, unless the right to the fox was established beyond all doubt by its actual capture.

In *Keeble* it was held that

> "When decoys have been used in order to take a profit for the owner of the pond and whereby the markets of the nation may be furnished; there is great reason to give encouragement thereunto" (Quoted in Berman and Hafner (1993)).

Here there is sufficient social value (an economic value) to make it worthwhile for the law to intervene. That the plaintiff was making his living makes the opinion that the activity is worthwhile beyond doubt.

When we come to Young with this background we can see that the decision must go for the defendant: the social utility is equal whether *Young* or *Hitchens* take the fish to market, and the land is unowned. To involve the law in such disputes would provide a fertile source of quarrels and actions without compensating social benefits.

Thus we can see that what might have led us to decide *Young* in the same way as *Keeble*, that in both cases the plaintiff was pursuing their livelihood, is not really relevant: the purpose of finding for the plaintiff in *Keeble* is neutralised in *Young* by the fact that the defendant is also pursuing their livelihood.

But note also that *Young* is not strengthened for the defendant by the fact that the plaintiff does not own the land: that is rather a separate argument, because it promotes a separate purpose: moreover it is not clear that ownership of a pond confers ownership of wild animals which temporarily alight on it.

Another case where two factors are both required is shown in Berman and Hafner (1993) in a discussion of Branting's Grebe system Branting (1991). Grebe is a system which models industrial compensation law through a semantic network representation of precedents. In Branting (1991) a case (decided for the plaintiff) where a person was injured when collecting iced water for workers at an oil drilling site, was held to support a decision in the favour of a schoolteacher collecting sandwiches for her lunch. Grebe, however, considers the argument weakened because the school is not unduly hot. This is obviously anomalous: the temperature is relevant in the first case because it establishes the operational need for iced water: the temperature of the school is irrelevant to establishing an operational need for sandwiches in the second case.

Understanding the teleology of the decisions thus enables us to steer away from two traps: seeing the common factor of *Young* and *Keeble* (pursuit of livelihood) as relevant when in fact it carries no weight when the defendant is similarly engaged: or as seeing the ownership of the land as important in *Keeble*, when in fact is a separate and rather weak argument.

6 Modelling Purpose

In the last section evidence was presented which shows that purposes motivating factors play an important role in reasoning with legal cases. If we represent legal cases in terms of factors which represent reasons for deciding for one side we should also represent the purpose which is promoted by deciding for that side in the presence of that factor. To return to our example, we could see the following purposes as relevant (other formulations are possible, but these will serve here):

- V1: Encouragement of economically valuable activity;
- V2: Clarity of the law;
- V3: Allowing the unfettered enjoyment of property to which one has title.

Now we can say for the factors in the example:

- If A the deciding for the plaintiff promotes V1;
- If B then deciding for the plaintiff promotes V3;
- If B then deciding for the defendant promotes V2 (since it is not clear whether a wild animal being on a person's land confers ownership);
- If C then deciding for the defendant promotes V2;
- If D then deciding for the plaintiff promotes V1;
- If E then deciding for the defendant promotes V3;

This allows us to distinguish cleanly between the case where two factors should be combined in a single argument from where the two factors represent different argument. If two factors promote the same purpose (for example both pursuing livelihood and having been awarded a medal for some activity would strengthen the argument that the activity is of value), they represent a single argument. If two factors promote different purposes they form separate arguments. Moreover we can now see why two arguments, each individually weaker than some third opposing argument, can be combined to defeat it. While the purpose promoted by the third argument may be valued more highly than either of the other two, the ability to promote two purposes at once may make it worth sacrificing the other purpose.

If we now return to the graph in figure one, but replace the factors by the purposes they promote, writing "deciding for the plaintiff promotes V1" as p -> V1, we get figure 2. I have included the purposes for the empty node on the plaintiff side in an effort to reflect the dissenting view in *Pierson*: fox hunting is possible of economic value, but holding this for an unpaid activity militates against clear law.

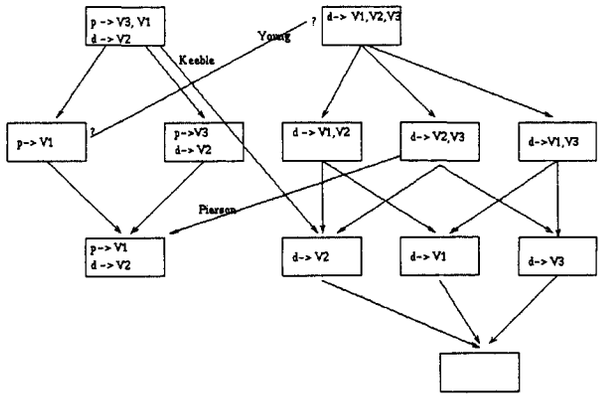

Figure 2: Purpose Orientated Representation of Pierson, Keeble and Young

We know explain the reasoning in the cases as follows. In *Pierson*, the decision should go to the defendant because the uncertainty surrounding possession of the animal, and the value of the activity, together with the desire to provide equal rights on common land, were considered (except by the dissenting judge) to outweigh the economic worth of the activity. In *Keeble* there can be no doubt that the activity is valuable, and this value is prized sufficiently to overcome the lack of clarity introduced. In *Young*, V1 points in neither direction: a finding for either side will promote V1. Once we remove the purpose moti-

vating factor A, it becomes clear that the case should be decided on the basis of *Pierson*.

Now, however, we reach this situation. Suppose we have linked our set of factors to the purposes they promote, and have a ranking which accords different weights to various purposes, and combinations of purposes. We can now deduce from these which side should be preferred in a given case. But we criticised accounts such as Prakken and Sartor (1998) for their inability to make other than arbitrary preferences between reasons not revealed in past decisions. Do we not simply move the problem a stage further back, and lay ourselves open to criticism because we cannot rationally defend the preference for one purpose over another? And if this is so, can we be said to be "completing the theory of demonstration by a theory of argumentation", where the latter is meant to be different in kind, rather than simply subject matter from the former? Is the ranking of purposes itself a subject of debate?

The answer seems to be "not explicitly". Judges are supposed in some way to embody the purposes which society wishes to promote and the preferences between them. As such they tend to act in a rather conservative manner, but are also open to change. Moles (1987) traces a shift in values in a series of cases involving the right to occupancy of a marital home in the context of domestic violence. From an initial position in which the purposes of protection of property were taken as paramount the situation shifts to one where the purposes of providing a violence free home for a woman to bring up her children in given preference. This series of cases generated a great deal of public debate, in which it became clear that the purposes promoted in the earlier decisions were no longer in accord with the public mood. This is often how the value preferences of judges shift.

The significance of this is that the things which are valued - and their relative values are removed from debate. The judges are supposed to have the values of society *as they currently are*, and to reason from this as a foundation. Judges are presumed to change their values as the society they represent changes its values: when they move too slowly public criticism is supposed to prompt the change.

As an aside one can see that this is why the prospect of a computer judge has always attracted such hostility. Applying the law (and the logic of past decisions) is one thing, and could well be done by a machine, but reflecting the values of a community requires that the life of the community be lived. This is an extreme case of the criticism often levelled at judges that their position removes them from the society the values of which they are supposed to reflect.

7 From Facts to Values: The Layers of Legal Reasoning

We are now ready to turn back to the initial question: do we need to complement a theory of demonstration with a theory of argumentation in order to model legal reasoning?

What we have seen is that decisions are made on particular cases, with a regard to the way past cases were decided (and - importantly - in the knowledge that they will be used to inform the decisions of future cases), so at to achieve certain purposes approved by society at large, and deriving from the judge's perception of the purposes. What seems mysterious is the move from facts to verdicts: often it seems that cases which are similar on the facts are treated differently, and that factually quite different cases are held to be similar. The move from facts to verdicts occurs in several stages. First the law evolves a series of distinctions (which are represented as factors in systems such as HYPO). Each distinction is made on the basis of facts in the case, but is not constrained by any common view of the facts. Consider the dog and bear example in section 5. Here the distinction between dogs and non-dogs is not zoological, but based on some perception of the typical properties of dogs which makes it right to classify bears as dogs and chihuahuas as non-dogs. The distinctions are introduced with the following in mind:

- If the distinction can be made in a case it should be a reason for deciding the case in a particular way - the distinction will favour either the plaintiff or the defendant, so that there is some point in making the distinction;
- The distinction should be justifiable in terms of some purposes which we (judges and society) wish to promote: dogs should be excluded to protect the safety, comfort and peace of mind of other passengers.
- The distinction should be capable of application, since the case is meant to stand as a precedent for future decisions.

One task for Artificial Intelligence would be to apply existing distinctions to categorise cases - or even draw new distinctions. No success has been reported here. Successful systems, such as HYPO Ashley (1990) and CATO Aleven (1997) represent - using the skill and judgement of the system builder - cases using the distinctions before applying their automated techniques. Systems which rely

on surface features of the case rather than developed distinctions used from analysis of past decisions seem to perform relatively poorly, as might be expected from the difficulty of automatically classifying cases according to these distinctions. What these systems can do is apply existing law when it determines an outcome (which is rare), or point to the distinctions favourable to each side which would suggest that that side should win. What typically happens is that the case is similar to some cases decided previously but the argument is weaker because some distinctions drawn in the previous case cannot be drawn in the new one. The decision therefore turns on whether the weakening of the argument is sufficient to turn the decision around.

If we want to discriminate between positions which are not determined by the distinctions alone, we need to consider the purposes promoted by the distinctions. Given a ranking on the comparative valuations of the purposes motivating distinctions, - which can itself be seen as being revealed in the past decisions - we can deduce which distinction should be held to carry the greater weight.

We thus have an interaction between various levels of consideration: facts; distinctions which can be made on the basis of facts; purposes motivating distinctions; and values attached to purposes. Within a given level demonstration is possible, but the terms that appear there - the distinctions which can made, for example - are justified by reference to their connection to terms at the next level.

8 Conclusion

We began by quoting Perelman's contention that the theory of demonstration need to be completed by a theory of argument. What our discussion has suggested is that this theory of argumentation perhaps comprises no more than an understanding of the layers that lie between fact and value, and the relationships between them. Against this background a theory of demonstration may still suffice. On this view reasoning with priorities completes the theory: dialectics are concerned with establishing the priorities. But the may be demonstrated from a suitable theory of value. In some comments on this work (Prakken 2000) Henry Prakken has shown how, by the addition of suitable axioms relating values to rules, an ordering on values, and a way of comparing the values promoted by rules, a formalisation of the situation with the logic of Prakken and Sartor (1998) is possible. Obviously the provision of such axioms will require some considerable work of knowledge representation. Never the less identifying the need for this knowledge, and the way in which it can be used, seems to point the way forward to some interesting and vital research.

References

Aleven, V., *Teaching Case Based Argumentation Through an Example and Models.* PhD Thesis. The University of Pittsburg. 1997.

Ashley, K.D., *Modelling Legal Argument.* Bradford Books, Cambridge, Mass, 1990.

Bench-Capon, T.J.M., *Some Observations on Modelling Case Based Reasoning with Formal Argument Models.* In Proceedings of the Seventh International Conference on AI and Law, Oslo, 1999.

Berman, D., and Hafner, C., Representing Teleological Structure in Legal Reasoning. In proceedings of the Fourth International Conference on AI and Law, Amsterdam, 1993.

Branting, L.K., Reasoning with Portions of Precedents, in Proceedings of the Third International Conference on AI and Law, Oxford, 1991.

Grasso, F., Cawley, A., and Jones, R.. Dialectical Argumentation to Solve Conflicts in Advice Giving. *International Journal of Human Computer Studies.* In press.

Kowalski, A., Case Based Reasoning and the Deep Structure Approach to Knowledge Representation. In Proceedings of the Third International Conference on AI and Law, Vancouver 1989.

Moles, R.N., *Definition and Rule in Legal Theory.* Blackwell, Oxford. 1987.

Perelman, C., *Justice, Law and Argument.* Reidel, Dordrect,. 1980.

Perelman, C., and Olbrechts-Tyteca, L., *The New Rhetoric*, Notre Dame Press, Notre Dame, 1969.

Prakken, H., (2000), *An Exercise in Formalising Teleological Reasoning.* Private Communication. Draft a Paper to be submitted to AI and Law.

Prakken, H., and Sartor, G., Modelling Reasoning with Precedents in a Formal Dialogue Game. *Artificial Intelligence and Law*, Vol 2-4, pp 231-87. 1998.

Reed, C., *The Role of Saliency in Generating Natural Language Arguments.* In Proceedings of the 16th International Joint Conference on Artificial Intelligence, 1999.

Twining, W., and Miers, D., *How To Do Things With Rules.* Weidenfeld and Nicholson, London. 1976.

Intelligent Systems to Support Deliberative Democracy in Environmental Regulation

Peter McBurney and Simon Parsons
Department of Computer Science, University of Liverpool
Chadwick Building, Liverpool L69 7ZF U.K.
{P.J.McBurney,S.D.Parsons}@csc.liv.ac.uk

Abstract

Among normative models for democracy, the Deliberative model suggests that public policy decisions should be made only following rational, public deliberation of alternative courses of action. We argue that such a model is particularly appropriate for the assessment of environmental and health risks of new substances and technologies, and the development of appropriate regulatory responses. To give operational effect to these ideas, we propose a dialectical argumentation formalism for an intelligent system within which deliberative debates about risk and regulation can be conducted. Our formalism draws on various philosophies of argumentation, scientific and moral discourse, and communicative action, due to Toulmin, Pera, Alexy and Habermas.

1 Environmental Regulation

New technologies and substances have the potential to cause adverse and unanticipated effects, on people, on other living species, and on our environment, and these effects are increasingly global in scale. Because of such risks, most chemicals and many technological innovations now require Government regulatory approvals before they can be used or sold commercially. In developed countries, there are now typically a number of Governmental and quasi-Governmental agencies tasked with assessing the risks of new substances and technologies and of framing appropriate regulatory responses for those perceived to have high adverse risks.

Framing and implementing legal regulations for such new substances is usually very difficult. Firstly, identification of all the potential consequences of a new substance can be problematic. Although thalidomide, for example, was safely tested on both animals and humans before its commercial release, none of the experimental subjects were pregnant, presumably because the possibility of differential harm was not considered (Teff & Munro, 1976). Recently, researchers in Artificial Intelligence (AI) have proposed the use of argumentation-based procedures for this problem of possibilistic risk assessment (Krause *et al.*, 1998; Fox, 1999).

Secondly, even when the possible consequences of new substances are believed known, assessment and quantification of risks is often problematic and invariably subjective (USA EPA, 1996; Shere, 1995; Rhomberg, 1997; Toll, 1999). In many cases, the scientific evidence upon which assessment is to be based is not conclusive and the theoretical issues involved are contentious, even among scientists working in the same field. The chemical formaldehyde, for instance, was found to cause statistically-significant increases in nasal cancers in rats but not in mice, while epidemiological studies of humans whose professions exposed them to high levels of the chemical found no significant increases in such cancers (Graham, Green, & Roberts, 1988). Moreover, quantification of risk normally requires the adoption of a mathematical model linking responses to exposure levels. Different dose-response models can result in widely different assessments of risk. Two theoretically-supported models for the risks associated with aflatoxin peanuts, for example, showed human risk likelihood differing by a factor of 40,000 (Pollak, 1996).

Another major issue for environmental and health regulation of new substances is that the consequences of different regulatory options may be very different. Page (1978) noted that different groups of people may benefit or lose from regulation or non-regulation of a substance, and that their gains or losses may be very different in scope, magnitude, intensity, timing and duration. A new chemical substance, for instance, wrongly deemed by regulators to be safe and then used, may adversely impact millions of people. How does one quantify the subsequent misery or loss of life? Conversely, the same chemical wrongly deemed to be unsafe, and so never sold, may cause large financial losses to the companies which undertook the initial research. Moreover, not using the chemical, when it would be safe to do so, may adversely impact those who could benefit from its use; these people too may number in their millions and their (unrelieved) misery may also be great.

Ultimately, regulation of new substances will always involve a trade-off of alternative regulatory options, weighing the anticipated consequences of each. As mentioned,

quantification of consequences is difficult. Yet, even without quantification of consequences, different values may be assigned to different consequences: society may, for instance, prefer to forego the sunk costs of commercial development of new chemicals rather than risk the loss of life resulting from their use. It is likely that different people and groups within society will have different value-assignments in such trade-offs. As an example, Stirling & Mayer (1999) assessed the different decision-criteria and value-assignments of a sample of British participants in the current debate over Genetically-Modified Organisms (GMOs). At present, however, there is no formal mechanism for articulating and comparing these valuations in most risk regulation; instead, different interest groups make their arguments and values known to regulators through lobbying and to the public through public relations activity. Keeney (1996) has argued for values to be assigned explicitly, so that hidden agendas may be exposed, and fairer and more transparent trade-offs undertaken.

By articulating and comparing the values of stakeholders, the multi-criteria scoring technique of Stirling & Mayer (1999) could be used in the formulation of regulatory policy. However, because this method is essentially quantitative and reductive, it does not incorporate the arguments for (or against) the criteria selected or the scores assigned. Our approach involves defining an intelligent system capable of qualitative representation and manipulation of arguments and values in the form of a dialogue game, and we have termed such a system a *Risk Agora* (McBurney & Parsons, 1999). Another approach, very similar in spirit to this, is the Zeno argumentation framework of Gordon & Karacapilidis (1997). This system was developed for conflict resolution and mediation in urban planning regulation, and uses an argumentation formalism based on the schema of Toulmin (1958) and the IBIS model of Rittel & Webber (1973). Our structure differs from Zeno not only in our intended application domain but also in our use of a specific philosophy of science to represent the community's understanding of scientific knowledge. This understanding may be distinct from that of any one participant, and needs to be formally represented. Another difference between our approach and that of Zeno is our deployment of dialogue moves based on speech acts specific to the domain of risk regulation, rather than more generic moves. Were environmental regulators to adopt an Agora framework for the development of risk regulatory policy, we believe this would give greater effect to a deliberative model of democracy, a concept explained in the next Section. Section 3 considers Deliberative procedures in the specific domain of risk regulation, and Section 4 outlines our formal structure for the Agora. Section 5 concludes.

2 Normative Models of Democracy

The term "Deliberative Democracy" was first introduced by Bessette (1980), and the subject has been the focus of much recent attention by philosophers of politics and law (Bohman & Rehg, 1997). It refers to a particular notion of democracy, one of several developed by philosophers as normative models of democratic politics.[1] To explain these, we begin with an abstract model of a democracy as consisting of just two entities: Society and the State. Society is the set of individuals, organizations and companies, together with the panoply of relationships between them, while the State is the apparatus of public-sector administration. The key normative question for democracy, then, is: *What should be the process of formation of political will?*, or *How should Society program the State?*

One model argues that a democracy is best governed when such programming is the task of a technocratic elite (who may be elected), making decisions on behalf of the general public. By contrast, rational-choice or liberal models view the process of political-will formation as akin to the workings of an economic market. In this model, political parties and interest groups act as entrepreneurs, offering alternative "products" in the form of bundles of state-instructions (or equivalently, philosophies of bundle- formation), to voters who then "purchase" their preferred bundle. That bundle with the greatest "market-share" — in the form of votes — becomes the set of instructions used to program the state.

The rational-choice model views citizens as consumers, acting in their own perceived individual self-interest and negotiating bargained compromises to political questions. Their preferences may be predetermined, uninfluenced by the process of choosing between alternative programmes. By contrast, deliberative notions of democracy see the political process as more than this, with citizens undertaking a substantial process of public deliberation to decide political questions. In this model, political-will formation is a process by which collective decisions regarding practical questions are made on the basis of rational and public reflection of the arguments for and against different courses of action. Such deliberation may well lead to the participants to change their preferences and their value assignments. As Michelman (1989) defined it:

"Deliberation ... refers to a certain attitude toward social cooperation, namely, that of openness to persuasion by reasons referring to the claims of others as well as one's own. The deliberative medium is a good faith exchange of views — including participants' reports of their own understanding of their respective vital interests — ... in which a vote, if any vote is taken, represents a pooling of judgments."

The benefits claimed for deliberative approaches to democracy include the legitimacy which public participation provides to political decision-making. People are

[1] Our presentation here draws primarily from Christiano (1997) and Chapter 9 of Habermas (1998).

generally more willing to accept decisions which they have had a role in forming, even when they disagree with the outcomes of those decisions. Moreover, the very act of participation may indicate, or may induce or strengthen, a concern for the welfare of the community beyond mere individual self-interest. In addition, a society with deliberative procedures may treat its citizens with more respect than it would with elitist or rational-choice procedures.

3 Deliberation in Risk Regulation

We believe that the specific characteristics of environmental risk regulation lead to further benefits from the adoption of public deliberative approaches. Firstly, the likelihood of identifying all possible consequences of new substances and technologies is increased the greater the number of participants engaged in considering the problem. Secondly, the inconclusive nature of much of the science involved and the subjective nature of risk assessments together mean that broad debate is invaluable. In this way, assumptions can be tested, experiments replicated, and inferences subject to detailed scrutiny. Many conclusions may fall when so subjected. For instance, Wynne (1996) argues that scientific experts often possess a generalized form of knowledge, which may not always be valid in particular circumstances. An example of this is shown by the case of alleged deforestation in West Africa (Fairhead & Leach, 1998), where the traditional inhabitants have understood local environmental reality better than western scientists.[2]

Thirdly, the complexity of most important risk assessments requires contributions from a wide spectrum of expertise. For instance, a rational decision regarding the regulation of Genetically-Modified Organisms (GMOs) arguably requires expertise in: molecular biology, genetics, plant biology, entomology, ecology, medicine, agriculture, economics, statistical experiment design, statistical inference, marketing, international trade, international development and international law. In such circumstances, as Willard (1990) has argued, no one person has the breadth of expertise required, and so even experts must perforce accept arguments based on authority from outside their own domain.[3] In this context, deliberative procedures can ensure assumptions from different disciplines are all valid and consistent, and explore cross-disciplinary interactions and conclusions. Finally, deliberative procedures can ensure the articulation of the consequences associated with different regulatory alternatives, and of the different values different people and groups may place on these. For all these reasons, we believe that deliberative procedures should ensure better quality decision-outcomes in environmental regulation than can either rational-choice or elite procedures.

How can deliberative procedures be implemented in a large, modern society, with possibly millions of citizens, thousands of elected representatives and perhaps thousands of environmental decisions in train at any time? Traditional means of public consultation, such as referenda and town-hall meetings, are arguably better suited to infrequent deliberations or to small communities. Citizens' Panels, where a small group of people, acting like a jury, hear evidence from relevant experts on an issue and then determine a course of action, have been used in some recent scientific policy debates.[4] Such panels, by exposing the participants to a diversity of expert opinion and forcing a decision to be made, are undoubtedly valuable for those involved; however, not everyone can participate. The Internet has been seen by some commentators (e.g. Ess (1996)) as a means to enable greater democratic participation in public policy decision-making, both through the wider availability of information and through electronic voting systems. We believe that a Risk Agora, suitably instantiated, could represent the scientific and political uncertainty involved in an environmental risk assessment and potentially bring the benefits of participation in a citizens panel to a wider audience. The next Section outlines the formal structure of our system.

4 Agora Formalization

In order that an intelligent system is able to be used for regulation decisions regarding a new substance (or technology), it will need to be able to do the following:

- Represent possible risks of deployment or non- deployment of the substance.
- Represent scientific uncertainty over the possibility, causal mechanisms, magnitude, duration and scope of potential risks.
- Represent different regulatory options for the substance and their potential consequences of enacting or not enacting these.
- Represent values assigned to such consequences by different individuals or groups.
- Enable the questioning, contestation, defence and qualification of each of the above types of statements.

[2] In coastal ecology, the importance of such local, "indigenous", knowledge for the selection of appropriate coastal-zone management policies has been recognized by the development of the *SimCoast* expert system by the U.K. Centre for Coastal Management (`www.ccms.ac.uk/simcoast.htm`). This has been used in several countries to aid coastal policy-making, by representing and integrating both scientific and traditional knowledges of local ecology. However, despite its incorporation of indigenous knowledge, representation in *SimCoast* is undertaken within a western scientific ontology, and assumes both that integration of different knowledges is achievable and that it is possible through discussion.

[3] U.S. company Applied Biomathematics is developing software to assess the validity of arguments across different disciplines in ecological risk assessments. See: `http://www.ramas.com/interest.htm#validity`.

[4] See Bhattachary (1998) for further description of such approaches.

- Enable the coherent comparison and manipulation of arguments for and against particular statements.
- Enable the synthesis of arguments into an overall case for a particular statement.
- Enable the taking of summary "snapshots" of a debate at any time.
- Enable the selection (and hence, legal imposition) of a particular regulatory option.

In (McBurney & Parsons, 2000), we developed a formal dialectical argumentation syntax, using a propositional language, for representing arguments over scientific claims of chemical carcinogenicity. This formalism drew upon, firstly, the philosophy of science of Pera (1994), which views scientific activity as a three-person dialogue game between a scientific investigator, Nature and a skeptical scientific community. Secondly, our structure drew upon rules for discourse in the philosophy of Discourse Ethics of Habermas (1991) and Alexy (1990), developed as a normative model for rational debate between reasonable and consenting participants. Although proposed initially for debates over moral questions, the theory has since been applied to legal and political domains, as in Habermas (1996). Thirdly, our formalism used argumentation schema of Toulmin (1958), within a dialectical framework, to enable the presentation of arguments for and against scientific claims. In other words, participants in the debate could variously posit, assert, contest, justify, rebut, undercut, qualify and retract claims, just as happens in real scientific discourse.

Moreover, by the use of dictionaries of uncertainty labels, our formalism permitted the assertion of individual degrees of belief in claims, their supporting evidences, their modes of inference, and their consequences. Participants could, for example, accept a scientific claim but label it as, say, *Plausible*, rather than as *Confirmed*. As an example of our formalism, a debate participant $\mathcal{P}_i$ could demonstrate her argument $\mathcal{A}(\rightarrow \theta)$ supporting a claim θ, an argument to which she was committed with strength D, by making the dialogue move:

$$\textbf{show\_arg}(\mathcal{P}_i : \mathcal{A}(\rightarrow \theta, D)).$$

By use of such dictionaries (which could be quantitative, e.g. probability estimates), degrees of commitment and uncertainty by individual participants can be represented. Our formalization also included truth-valuation functions which assigned degrees of certainty to statements on behalf of the community as a whole, based upon the existence and strength of arguments for and against the statement and its supporting grounds. Such truth-valuation functions effectively produce an on-going representation of the dialogue community's changing views of a scientific claim, and so provide the desired "snapshot" capability.

However, as the discussion in this paper has demonstrated, in a regulatory context there are other types of statements besides scientific claims: statements of values and preferences, moral obligations and relationships, and imperatives (i.e. regulations). Therefore, in this paper, we extend our earlier formalism by incorporating expressions for these additional types of statements. We do this by drawing on other work of Habermas, his philosophy of Communicative Action (Habermas, 1984), in which he sought to understand how people collaborate rationally to achieve a common understanding of a situation or a collective action. As part of his philosophy, he proposed a typology of statements,[5] which we have adapted and relabelled for the specific context of environmental regulation. (Habermas' labels are in parantheses.)

Factual Statements (Constative Speech Acts): These are statements which seek to represent the state of the external world. In our domain, such statements include claims about scientific reality, and the scientific, economic or social consequences of particular actions. In the Agora formalism, we demarcate these different types of factual statements for clarity of exposition. Contesting such a statement means denying that it is a true description of objective, external reality.

Value Statements (Expressive Acts): These are statements which seek to represent the state of the speaker's internal world, i.e. they reveal publicly a subjective preference or value assignments. Such statements may only be contested by doubting the sincerity of the speaker.

Connection Statements (Regulative Acts): These are statements which assert some relationship between different parties, in the common world of the Agora participants. One may assert, for example, that the stakeholders with an interest in the regulation of a proposed new technology is wider than previously defined.

Inferential Statements (Operative Acts): These are statements which refer to the content of earlier statements made in a debate, drawing inferences from them or assessing implications. Once a scientific theory has been proposed, a specific risk assessment model and the ensuing calculations based on this theory fall into this category. Contestation of such statements can take the form of questioning the appropriateness or the validity of the inferences made.[6]

Procedural Statements (Communicative Acts): These are statements about the activity of speaking itself, such as the rules for participation and debate. In many real-life discourses, these often become the focus of debate, overtaking issues of substance.[7]

[5] He was building on the typology of Searle (1979).

[6] Our definition departs slightly from that of Habermas, in that our Inferential Statements may have "genuine communicative intent."

[7] For instance, in the scientific debate over GMOs in Britain during

Obligation Statements (Imperative Acts): These are statements which assert some obligation on the participants, for example, that they must limit the commercial sale of a new substance. Only the authorized regulator has the power to make such assertions, and once made, cannot be contested within the Agora. (In real-life, they may of course be contested in the courts.)

Given this typology of statements possible within the Agora, we can define a syntax of dialogue moves, extending the syntax for scientific reasoning of McBurney & Parsons (2000). Thus, for instance, a debate participant $\mathcal{P}_i$ could state her value assignment D_C to consequence C of action θ by means of the dialogue move:

show_value$(\mathcal{P}_i : Val(\theta \rightarrow C, D_C))$.

Debate over such statements will then proceed according to the same rules for positing, proposing, contesting, qualifying, etc, statements as defined in our earlier paper.

The dialectical argumentation formalism we have presented here is related to other recent work in which we have applied argumentation in the design of intelligent systems. For example, in Fox & Parsons (1998); Parsons & Green (1999), we developed formalisms for the articulation and manipulation of statements of qualitative value, as part of calculi for qualitative decision-making. In both these papers, the argumentation formalism presented was monolectical, whereas in Amgoud, Parsons, & Maudet (2000), we presented a formalism for dialectical argumentation, involving two participants engaged in a generic debate.

5 Example

In this section, we illustrate our approach with a simplified example drawing on recent debates over Genetically Modified Organisms (Stirling & Mayer, 1999; UK ESRC Global Environmental Change Programme, 1999). We assume a debate with willing and reasonable participants denoted $\mathcal{P}_1, \ldots, \mathcal{P}_6$. For ease of understanding, we articulate the dialogue moves in plain English (not in the formal syntax), and we label each move with its type.

M1 (Factual): $\mathcal{P}_1$ asserts that foods containing GMOs may not be safe to eat.

M2 (Query): $\mathcal{P}_2$ asks $\mathcal{P}_1$ for an argument supporting Claim M1.

M3 (Factual): $\mathcal{P}_1$ presents evidence of experiments in which rodents fed GM potatoes had significantly greater tumors than a control group.

1999, an argument between the medical journal *The Lancet* and The Royal Society ensued over whether the latter was entitled to comment on a paper submitted to the journal while it was still under consideration for publication (Lancet Editorial, 29 May 1999).

M4 (Value): $\mathcal{P}_3$ accepts the assertion M1 of $\mathcal{P}_1$, and places a large negative value on GMOs being in the human food chain.

M5 (Inferential): $\mathcal{P}_2$ asserts that assertion M1 of $\mathcal{P}_1$ only follows from argument M3 if humans and rodents are sufficiently alike biochemically.

M6 (Value): $\mathcal{P}_4$ places zero value on GMOs being in the human food chain, provided they are labeled whenever present.

M7 (Connection): $\mathcal{P}_5$ says asserts that agriculture and food distribution companies have a duty to inform consumers of the presence of GMOs in their food products.

M8 (Inference): $\mathcal{P}_2$ asserts that labeling of foods containing GMOs (as suggested in moves M6 of $\mathcal{P}_2$ and M7 of $\mathcal{P}_5$) will lead to consumers rejecting such foods in favor of non-GM foods, even if they pose no dangers.

M9 (Factual): $\mathcal{P}_6$ asserts that GM foods have the potential to end hunger in the Third World.

M10 (Query): $\mathcal{P}_5$ asks $\mathcal{P}_6$ for evidence for her assertion M9.

M11 (Factual): $\mathcal{P}_6$ provides evidence of the increased yields available from GM crops.

M12 (Contestation): $\mathcal{P}_3$ contests assertion M9 of $\mathcal{P}_6$, on the grounds that the high investments required for development of GM Foods will lead to increasing concentration of corporate ownership in the agrichemicals sector, and this will increase poverty in the developing world.

M13 (Contestation): $\mathcal{P}_6$ defends assertion M9 against the attack of $\mathcal{P}_3$ by contesting the claim that increased concentration of corporate ownership will necessarily increase poverty in the developing world.

M14 (Procedural): $\mathcal{P}_5$ asserts that this debate should be limited to a discussion of the consequences of permitting GMOs only in our own country, and not elsewhere.

M15 (Value): $\mathcal{P}_4$ asserts that she would place greater value on a regulatory decision made with awareness of the global consequences than one informed only by local consequences.

... and so on.

As this very simplified example shows, the scientific, economic and social issues involved may be quite complex, and the values placed on outcomes by different participants possibly very discordant. There is no guarantee of resolution of such differences, as Stirling & Mayer

(1999) found in their application of multi-criteria scoring to the same issue of GM foods. However, even without a guarantee of resolution, representation of a debate within such a formalism, forces greater clarity in the statements articulated, and this will surely facilitate any attempt at trade-offs between different alternatives.

6 Discussion

In this paper, we have proposed a novel dialectical argumentation formalism for an intelligent system within which deliberative debates about possible environmental risks and regulatory alternatives can be conducted. As we have explained, our formalism draws on Toulmin's theory of argumentation, Pera's philosophy of science, the Discourse Ethics of Habermas and Alexy, and Habermas' theory of Communicative Action. We see this approach as potentially giving practical effect to notions of Deliberative Democracy, enabling rational, public and transparent consideration of decision alternatives prior to deciding on a course of action. As well as effecting deliberative notions of democracy, this approach, we have argued, provides particular benefit in the domain of regulation of environmental and health risks. This is due, firstly, to the typically difficult, subjective and contested nature of risk assessment and the science on which it is based. Secondly, it is because regulatory alternatives may impact different groups with markedly different consequential outcomes, outcomes to which individuals and groups may assign very different values. Once instantiated with the details of a specific risk debate, the Risk Agora could be used in a number of ways:

1. To understand the logical implications of the scientific knowledge relating to the particular issue, and the arguments concerning the consequences and value-assignments of alternative regulatory options.

2. To consider the various arguments for and against a particular claim (including regulatory options), how these arguments relate to each other, their respective degrees of certainty, and their relative strengths and weaknesses.

3. To develop an overall case for a claim, combining all the arguments for it and against it.

4. To enable interested members of the public to gain an overview of the debate on an issue.

5. To support group deliberation on the issue, for example in Citizens Panels.

6. To support risk assessment and regulatory determination by government regulatory agencies.

As the last three in this list demonstrate, the Risk Agora potentially gives effect to the ideals of reasoned, public decision-making, and thus supports notions of deliberative democracy. We believe the nature of decisions involved in the assessment and regulation of risk mean that the adoption of such processes may improve the quality and fairness of decisions made in this domain.

Acknowledgments

We are grateful for partial funding from the British EPSRC, under grant GR/L84117 and a PhD studentship. We also thank William Rehg and Trevor Bench-Capon for discussions on these topics.

References

Alexy, R. 1990. A theory of practical discourse. In Benhabib, S., and Dallmayr, F., eds., *The Communicative Ethics Controversy*. Cambridge, MA, USA: MIT Press. 151–190. (Published in German 1978).

Amgoud, L.; Parsons, S.; and Maudet, N. 2000. Modelling dialogues using argumentation. In *Proceedings of the International Conference on Multi-Agent Systems, Boston, MA, USA*.

Bessette, J. 1980. Deliberative Democracy: The majority principle in republican government. In Goldwin, R. A., and Schambra, W. A., eds., *How Democratic is the Constitution*. Washington, DC, USA: AEI. 102–116.

Bhattachary, D. 1998. *Enviromental values: a critical appraisal of cost-benefit analysis and discursive forms of decision-making*. Report, Oxford Centre for Environment, Ethics and Society, Mansfield College, Oxford, UK.

Bohman, J., and Rehg, W., eds. 1997. *Deliberative Democracy: Essays on Reason and Politics*. Cambridge, MA, USA: MIT Press.

Christiano, T. 1997. The significance of public deliberation. In Bohman, J., and Rehg, W., eds., *Deliberative Democracy: Essays on Reason and Politics*. Cambridge, MA, USA: MIT Press. 243–277.

Ess, C. 1996. The political computer: Democracy, CMC, and Habermas. In Ess, C., ed., *Philosophical Perspectives on Computer-Mediated Communication*. Albany, NY, USA: State University of New York Press. 197–230.

Fairhead, J., and Leach, M. 1998. *Reframing Deforestation: Global Analysis and Local Realities: Studies in West Africa*. London, UK: Routledge.

Fox, J., and Parsons, S. 1998. Arguing about beliefs and actions. In Hunter, A., and Parsons, S., eds., *Applications of Uncertainty Formalisms*. Berlin, Germany: Springer Verlag (LNAI 1455). 266–302.

Fox, J. 1999. Will it happen? Can it happen? A new approach to formal risk analysis. *Risk, Decision and Policy* 4.

Gordon, T. F., and Karacapilidis, N. 1997. The Zeno argumentation framework. In *Proceedings of the 6th International Conference on AI and Law*, 10–18. ACM Press.

Graham, J. D.; Green, L. C.; and Roberts, M. J. 1988. *In Search of Safety: Chemicals and Cancer Risk.* Cambridge, MA, USA: Harvard University Press.

Habermas, J. 1984. *The Theory of Communicative Action: Volume 1: Reason and the Rationalization of Society.* London, UK: Heinemann. (Published in German 1981).

Habermas, J. 1991. *Moral Consciousness and Communicative Action.* Cambridge, MA, USA: MIT Press. (Published in German 1983).

Habermas, J. 1996. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy.* Cambridge, MA, USA: MIT Press. (Published in German 1992).

Habermas, J. 1998. *The Inclusion of the Other: Studies in Political Theory.* Cambridge, MA, USA: MIT Press. Edited by C. Cronin and P. De Greiff.

Keeney, R. L. 1996. The role of values in risk management. *Annals of the American Academy of Political and Social Science* 545.

Krause, P.; Fox, J.; Judson, P.; and Patel, M. 1998. Qualitative risk assessment fulfils a need. In Hunter, A., and Parsons, S., eds., *Applications of Uncertainty Formalisms.* Berlin, Germany: Springer Verlag (LNAI 1455). 138–156.

Lancet Editorial. 29 May 1999. Health risks of genetically modified foods. *The Lancet* 353(9167).

McBurney, P., and Parsons, S. 1999. Truth or consequences: using argumentation to reason about risk. *Symposium on Practical Reasoning, British Psychological Society, London, UK.*

McBurney, P., and Parsons, S. 2000. Risk Agoras: dialectical argumentation for scientific reasoning. *In submission.*

Michelman, F. I. 1989. Conceptions of democracy in American Constitutional argument: the case of pornography regulation. *Tennessee Law Review* 291:293.

Page, T. 1978. A generic view of toxic chemicals and similar risks. *Ecology Law Quarterly* 7 (2):207–244.

Parsons, S., and Green, S. 1999. Argumentation and qualitative decision making. In Hunter, A., and Parsons, S., eds., *ECSQARU99: The 5th European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty.* Berlin, Germany: Springer Verlag (LNAI 1638). 328–339.

Pera, M. 1994. *The Discourses of Science.* Chicago, IL, USA: University of Chicago Press.

Pollak, R. A. 1996. Government risk regulation. *Annals of the American Academy of Political and Social Science* 545:25–34.

Rhomberg, L. R. 1997. A survey of methods for chemical health risk assessment among Federal regulatory agencies. *Human and Ecological Risk Assessment* 3 (6):1029–1196.

Rittel, H. W. J., and Webber, M. M. 1973. Dilemmas in a general theory of planning. *Policy Sciences* 155–169.

Searle, J. 1979. *Expression and Meaning: Studies in the Theory of Speech Acts.* Cambridge UK: Cambridge University Press.

Shere, M. E. 1995. The myth of meaningful environmental risk assessment. *Harvard Environmental Law Review* 19 (2):409–492.

Stirling, A., and Mayer, S. 1999. *Rethinking Risk: A Pilot Multi-Criteria Mapping of a Genetically Modified Crop in Agricultural Systems in the UK.* Report, SPRU, University of Sussex, Brighton, UK.

Teff, H., and Munro, C. R. 1976. *Thalidomide: The Legal Aftermath.* Westmead, Farnborough, Hampshire, UK: Saxon House.

Toll, J. E. 1999. Elements of environmental problem-solving. *Human and Ecological Risk Assessment* 5(2):275–280.

Toulmin, S. E. 1958. *The Uses of Argument.* Cambridge, UK: Cambridge University Press.

UK ESRC Global Environmental Change Programme. 1999. *The Politics of GM Food: Risk, Science and Public Trust.* Special Briefing No. 5. Brighton, UK: University of Sussex.

USA EPA. 1996. Environmental Protection Agency: Proposed guidelines for carcinogen risk assessment. *U.S. Federal Register* 61:17960–18011 (23 April 1996).

Willard, C. A. 1990. Authority. *Informal Logic* 12:11–22.

Wynne, B. 1996. May the sheep safely graze? A reflexive view of the expert-lay knowledge divide. In Lash, S.; Szerszynski, B.; and Wynne, B., eds., *Risk, Environment and Modernity: Towards a New Ecology.* London, UK: Sage. 44–83.

THE ELECTRONIC JUDGE

Pedro Valls Feu Rosa

JUSTICE
STATE SUPREME COURT OF APPEALS OF ESPIRITO SANTO
BRAZIL
pvfr@zaz.com.br

Abstract

Let us take a look at the judges' daily routine. They are, usually, responsible for hundreds, even thousands of cases, some of high complexity, involving human problems, demanding more than logical interpretations; these cases d e-mand judges with "blood in their veins". However, quite a number of them are related to trivial questions, repetitive su b-jects in which there are no doubts to which would be the fair decision.

1 Introduction

In 1958, the powerful President of IBM Corporation, Mr. Thomas J. Watson, said that "there is a world ma r-ket for about five computers, and no more".

Lord Kelvin, the famous physician and mathematician, once proclaimed that "flying machines heavier than the air are impossible to work".

And why not also quote Wilbur Wright, one of the pi o-neers of aviation? In 1901, after experiencing a failure with his airplane, he said: "man won't fly in the next thousand years".

Auguste Lumière, in his turn, proclaimed that "my i n-vention (the projecting camera) is not to be sold. It will be used for some time as an amusement, but it doesn't have any commercial future".

Not less amazing were the words of Mary Somerville, the pioneer of radio broadcast, when stated, in 1938, that "the television won't last; it is a fashion that will be short-lived".

A more recent thought came from an Admiral Willian D. Lahy's statement to President Truman, in 1945: "That was the greatest foolishness we have ever done. The atomic bomb will never explode, and I speak as a sp e-cialist in explosives".

Equally unfortunate was the American magazine "The Literary Digest" about automobiles in 1899: "The called "horseless carriage" is presently a luxury for the rich, and although its price should be reduced in the future, it will never get to be of common use, as the bicycle is".

I am quoting only smart, intelligent and well educated people, leaders of their time.

Those reflections came to my memory when I read an article written in the beginning of the century about the introduction of the typewriters in the Judgement Courts. In that acid article it was emphasized that the typewriter would be an extremely harmful innovation to the ser v-ices of Justice because it would facilitate frauds, in a way that it would be impossible to know with certainty who wrote a decision. Today it's clear the stupidity of those points, if sustained to the time to a learned magistrate, conscious of his authority.

In fact we are facing another technological revolution: the computer revolution. And our behavior has been similar to those above-mentioned skeptical words: we are using the computers as mere "show-off typewriters" or as a sophisticated card index.

I agree with the French journalist Émile de Girardin, when he says that "everybody speaks of progress, but nobody leaves the routine". I'm sorry to say that a wo n-derful invention like computers still is far away from the present reality of our legal system...

I believe that it is important to point out that if researc hers on artificial intelligence didn't make "thinking m achines", they made possible the development of softwares that simulate aspects of human reasoning, solving problems with uncommon efficiency - the so called "sp ecialist systems", widely used in medicine, business, aviation and so on. Nevertheless those systems resist in being used in the world of the law.

2 The Electronic Judge

So let us take a look at the judges' daily routine. They are, usually, responsible for hundreds, even thousands of cases, some of high complexity, involving human pro blems, demanding more than logical interpretations; these cases demand judges with "blood in their veins". Ho wever, quite a number of them are related to trivial que stions, repetitive subjects in which there are no doubts to which would be the fair decision.

Let us imagine, for instance, the majority of traffic acc idents. Any person, although nonprofessional, with some good sense, is capable of defining the person's respons ibilities in most of the cases. The rules to be followed are usually so simple, that a bit of logic is usually enough in order to get a decision.

Of course we could argue that there is always, even in simple traffic case decisions, a human factor to be co mplied with. I dare to state that this factor really doesn't exist, as in many other cases examined not only by judges, but for judges and the public administration, in general.

In spite of any emotion, the fact is that judges, in their decisions, must register exactly all the elements that weigh in their convictions - and those elements are purely logical. I propose, as an example, a sort of a tour around the main aspects that include the judgement of a traffic accident (the example that we used now). More over, I suggest an example through a service called "Ju stice on Wheels", at my hometown, an idea I had that is real and in practice.

Let us think about a simple case that happens ever ywhere in the world: in a crossing, two cars crashed. The drivers arguing with each other on whoever was right at the correct traffic light color. Within a few minutes, the "Justice on Wheels" vehicle gets to the place of the crash with a Judge and his/her team on board. Let us think about what will be seen and what will be considered to be done in terms of sentence.

The first evidences of course will be taken from the crashed vehicles: the damages, possible tire marks, place characteristics, etc., which is recorded by the police off icer and rewritten word-for-word by the judge, in order to prepare the decision. Those procedures take time. So why not insert all the information directly on the co mputer? I will now state, for your benefit, some advantages of such a procedure:

First of all, one single data will be inserted on the co mputer, which avoids the retyping of standard wording.

Secondly, the software will evaluate, with more accuracy than a human being ever could, the effects of the veh icles speed and of ground conditions at the accident.

After examining the place of the crash the magistrate shall hear the witnesses and certainly get some inform ation on the drivers conduct, exclusively under an obje ctive approach. He/she will consider, also based on obje ctive elements, whether or not the witnesses are reliable. All of the components of the accident will be written in full detail in order to be used at the making of the final decision. I even dare to add, at this point, that all the writing work will be mechanical. There's no doubt about that, for, after all, all the basic elements of the case are unquestionable. For example:

1. Did the driver stop at the red light? Yes or no?

2. Was the driver speeding, considering the ground and weather conditions? Yes or no? What was the approx imate speed?

3. Had the driver been drinking alcohol above the a cceptable limit of the law? Yes or no?

4. Did the witnesses have any sort of connection with the drivers (friendship, for example)? Yes or no? If they did, of what kind was their relationship?

All of the questions above, and I couldn t stress this enough, are purely objective. There are no answers to them other than a yes or no.

With this information at hand, the judge will write the final sentence, which is something that normally takes a great deal of time. So, why not write down the answers directly on the computer? The work of the magistrate will be limited to making the choice between some o p-

tions that will come up on the screen and subsequently the system by itself will make the whole report of the facts.

Then comes the great moment, in which the judge will make the analysis of the elements in the scene. However, the analysis will not be made under any subjective point of view. There is indeed no way in which that could happen! The judge will have to write in full detail all the reasons that shaped his decision. So in that kind of work there is no gap for a subjective sentence. The decision will be based only on objective elements of the case; those being, in the example given , the vehicles, the wi t-nesses and so forth.

I emphasize that the judge cannot make remarks such as "I feel that the blame belongs to Paul", or "I think that Mary caused the crash". In the same way as the m a-chine, the judge is confined to the logical characteristics of the case.

And so I ask: if we are concerned with nothing more than pure logic, why not to give the task of the decision to a good computer software? We would save time and achieve more detailed conclusions. It is important that we do not forget how modern life has increased the fr e-quency in which judges give shortened and imperfect decisions. And, not only has modern life been doing that, it keeps on doing so more and more each day as time goes by.

The skeptical one would question the possibility of whether a software would be reliable enough to do that kind of work meant for a judge only. The answer is yes - I developed a computer program that elaborates judicial decisions. All on it's own, it comes to conclusions and composes texts, depending on nothing beyond the use of a "mouse", which will mark the options shown on the screen.

I am not speaking about a mere "decision tree" (I tried this idea, but I failed), nor of using " standardized" texts (the possibilities are practically infinite). I developed a program that composes the final decision.

I would also like to point out that the version of the pr o-gram I use at my office can actually "read" the report of a case and "understand" what is being said. It then "searches" for the best solution and goes on to compose, automatically, the decision and it does all this without my interfering in it at all. I only ever interfere when the interference is aimed at "teaching" it and increasing it s efficiency for future cases.

This software and the idea in it, copyrighted under the name of "The Electronic Judge", can be used to elaborate practically any judicial decision, in public and private institutions, all over the world - it is only a matter of making new versions of it!

On traffic accidents, this software is succeeding at a 68% level of acceptance. In my office, at the Appeals Court, I use another version of this software, with great success. And the more accurate the software, added to the pra c-tice it gets as it is used in the daily routine, the better the decisions will, obviously, come out.

People will ask: is a judge being replaced by a machine? That's, clearly, far from being the idea. The software's decisions can be freely changed by human beings and even modified by them. Furthermore the judge will sign the sentence just in case he/she agrees with it, and if it is a correct decision. Thus the machine doesn't replace the judge; it is only limited to doing the logical and m e-chanical work for him.

In fact, for today's standards, we can not talk of such a possibility. Computers operate hospitals and hotels, they land planes with hundreds of lives on board, they even decide without any assistance our commercial and banking records. These and an endless number of other things. Well, after all, the required mental work to judge, for instance, a non-victim car crash, isn t very large. And so, the problem, on one hand, is the analysis of a pile of identical cases and, on the other, thousands of low complexity incidents that afflict Justice and co m-mon people, all over the world, people who deserve, as taxpayers, better attention. Going back to the subject of the "Electronic Judge", we would adopt and enlarge the resources that Computer Science gives us all, saving the judges' time so that they would be able to dedicate more of it to those cases with a higher degree of complexity, where it will be necessary to make efforts beyond the reach of simple logic applications. The same goes to public administration, the private companies, and all the places that a judicial decision is needed.

I am not at all in favor of replacing educated decision makers by machines. However, I strongly defend the replacement of mental routine by electronic routine, giving people more time to evolve as human beings, at the same time that the efficiency of society is being i m-proved.

The Jama Legal Narrative. Part II: A Foray Into Concepts of Improbability

Ephraim Nissan

School of Computing and Mathematical Sciences. University of Greenwich,
Wellington Street, Woolwich, London SE18 6PF, U.K.
E.Nissan@greenwich.ac.uk http://www.gre.ac.uk/~E.Nissan
http://www.gre.ac.uk/~E.Nissan/CmsWeb/ai_and_law_group.html

Abstract

In judicial decision making, the doctrine of chances takes explicitly into account the odds. There is more to forensic statistics, as well as various probabilistic approaches which taken together form the object of an enduring controversy in the scholarship of legal evidence. In this paper, we reconsider the circumstances of the Jama murder and inquiry (dealt with in Part I of this paper: "The Jama Model. On Legal Narratives and Interpretation Patterns"), to illustrate yet another kind of probability or improbability. What is improbable about the Jama story, is actually a given, which contributes in terms of dramatic underlining. In literary theory, concepts of narratives being probable or improbable date back from the eighteenth century, when both prescientific and scientific probability was infiltrating several domains, including law. An understanding of such a backdrop throughout the history of ideas is, I claim, necessary for AI researchers who may be tempted to apply statistical methods to legal evidence. The debate for or against probability (and especially bayesian probability) in accounts of evidence has been flourishing among legal scholars; nowadays both the Bayesians (e.g., Peter Tillers) and the Bayesio-skeptics (e.g., Ron Allen) among those legal scholars who are involved in the controversy are willing to give AI researchers a chance to prove itself and strive towards models of plausibility that would go beyond probability as narrowly meant. This debate within law, in turn, has illustrious precedents: take Voltaire, he was critical of the application of probability even to litigation in civil cases; take Boole, he was a starry-eyed believer in probability applications to judicial decision making (Rosoni 1995). Not unlike Boole, the founding father of computing, nowadays computer scientists approaching the field may happen to do so without full awareness of the pitfalls. Hence, the usefulness of the conceptual landscape I sketch here.

1 "Not Likely", yet "True"

In Part I, the story was told and discussed of a hate crime: a Somali refugee was burned to death as he was sleeping on a bench in a public park. Eyewitnesses identified the suspects, who were nevertheless released. What is peculiar about the eyewitnesses, is that they were football referees and linesmen eating at a restaurant close by. Here in Part II, we are going to ponder about what it is that makes this narrative so awkward in addition to its being strikingly saddening.

What is it that makes this story stand apart? More specifically—setting aside the narrative aspects that in Part I have been discussed in relation to concepts from automated narrative understanding within natural-language processing research—what is it about the kind of coincidence affecting the testimony, which makes the Jama story stand apart from other crime narratives which through the lens of investigation and prosecution also involve reasoning about the odds?

In Part I, this statement was made:

> [T]he eyewitnesses involved [...] happen to be football referees or linesmen: could you ever get more eagle-eyed an eyewitness that such ones?[1] Not that the average club fan is above insulting the referee when the referee does not rule in favour of his team. Yet, out of the stadium, that the eyewitnesses to a clamorous crime happened to be a group of such "professional witnesses" as it were, this *does* sound like it is taken from fiction. These eyewitness, interrogated during the inquiry, concorded in identifying a few persons [ldots] as being the perpetrators of the murder.

The approach was mentioned in Part I, by which the architecture of the JAMA prototype has a module envisaged whose input is the plot, and whose output is some remark about the narrative as though it was fiction. In such a perspective, we discussed likening devices in the plot to the manner of Bertolt Brecht, a playwright "known to resort to extreme elements in the narrative for dramatic underlining while making a claim whose truth he was purporting for real society (as per criteria he had laid down for his 'epic theatre')".

It need not be just a given author or a given genre; for other cultural contexts other parameters may more relevantly apply. Clinton Bond (1994) discusses realism in the early English novel, with the intent "to discover what the unspoken and elusive claims of realistic fictions were" (p. 122). "The novel's claim to be real, even—perhaps, particularly—while recognized as fiction, lies at the very heart of the genre, and should be seen less as a bizarre attempt to pass fiction as true than as a characteristic strategy built on the assertion that novels occupy exactly the same world–ideological and concrete—as their readers" (p. 123). Part I stated: "What I am claiming here is that not just narrative schemata about interpersonal or social situations, but specifically literary models as well may to some extent affect the way we ordinarily make sense of the world—including lay perceptions of legal narratives".[2]

We are going to see how the history of ideas about probability (even mathematical probabilities) happened to meet literary theory in a period that valued poisedness and classicistic harmony: eighteenth-century critique, by Lennox, of Shakespeare's Baroque excessiveness verged on its being detrimental to narrative probability. Come to that, in my assessment (Nissan 1996) of why some exculpatory explanations generated by my ALIBI programme (see Part I, end) seem improbable, I pointed out that this happens when the input accusation includes such damning evidence that by trying too hard to find an excuse, the resulting narrative becomes improbably contrived: see the next section.

In the AI literature, a naive passing reference to character evidence in relation to likelihood for the purposes of narrative understanding in terms of ascribed intentions occurs in the following. Calistri (1990), who proposes a probabilistic approach to misconception-detection in plan recognition, shortly relates the latter field to user models *(ibid.:* Section 6.6.1):

> [E]ven perfect plan recognition is frequently ambiguous. As an observer traces the first few actions of an agent, or as a listener hears the first few sentences of an utterance, the complete plan that is being followed is often ambiguous [Goldman and Charniak 1988]. For example, if we were to hear the sentence "Jack went to the supermarket", we could come up with many possible plans that Jack could be following: he could be going there to do some shopping, he could be going there to meet someone, he could be going there to rob the supermarket, or he may just be out for a stroll.

[1] In "Adversarial and inquisitorial proceedings", a comparative overview, University of Keele's Jenny McEwan remarks: "Credibility is, in adversarial systems, an issue of only collateral importance. To the judicial mind, judging the veracity of others is an assessment jurors are accustomed to make in their daily lives. Their common sense and experience make them ideally suited to the task, which they should perform unaided. Experts have no business telling juries what to think, unless they have a claim to superior judgement. On matters of credibility, such a claim exists only where the witness has a personality or has characteristics (such as being a child) which take the issue beyond the everyday experience of jurors" (1995: p. 503).

Psychiatrists are a prominent kind of expert witnesses; moreover, clinicians in their practice also take decisions that immediately affect a patient's status. *The Clinician as de facto Magistrate* is the subtitle of a book in forensic psychiatry (Bluestone et al. 1994); that subtitle "gives implicit recognition to the fact that clinicians in psychiatric practice today wield great influence over the law as it affects their patients' lives" (Lonsdorf 1995: p. 319).

In a sense, a football referee may wield almost supreme power (subject only to intervention of the police) over the universe of football matches, yet even in a situation as in the Jama story, the lynx-eyed football referee *qua* eyewitness would perhaps not wield in the courtroom more that (or as much as) the far from undisputed impact the forensic psychiatrist *qua* expert witness could have on a trial.

[2] In one of the final footnotes of Part I, I introduced *epistemic metaproperties* (Nissan 1987, 1995) to account for typicalness or even excess (a koala or a panda is "more a [teddy] bear than a bear" and the like); I related this to disproportionateness or excessiveness in the circumstances of the Jama case: a football referee, and all the more so a group of football referees and linesmen, is a "disproportionate", excessive, too-good-to-be-true; along with the release of the suspects ("These are good young fellows"), such excessive trait of the testimony makes, by contrast, for a grotesque situation. In the arts (of which literary expression is but a kind), the excessive is associated with the Baroque and generalised baroque (anticlassical) trends. In particular, this includes for example the narratives of Romanticism. Not by chance, Vols. 1 and 2 of Morse's *American Romanticism* (1987) were given the subtitles *Excessive America* and *The Enduring Excessive,* in that order. In the narrative we discussed, though, especially if one is to insist on the Brecht reminiscence, Expressionist art may be recalled by way of association, because a brutal impact is sought there on purpose.

> In most cases, however, we would immediately jump to the conclusion that Jack is going to the supermarket to do his shopping, and would not even consider the hundreds of other possible explanations. This sort of reasoning can be explained by using the information provided by a user model. If we aren't given any evidence to the contrary, we would assign Jack to some generic user model, which tells us that most people go to the store to do shopping. If we have heard that Jack is a known fugitive on the FBI's most wanted list, we might assign him to a "dangerous-felon" user model, which would prefer the explanation of robbing the store. Different user models will have different preferences for plans in the domain.

Such a passage is far from unproblematic in a legal perspective. Suppose an AI tool is developed, with a general part—which incorporates some standard technique based on the above in one of its modules—and a specialised part of its architecture, which is devised for legal purposes. The inadequacy of the idea quoted, *vis-à-vis* legal principles and in particular the U.S. exclusionary rules of evidence in court,[3] could provide an argument against the validity of the outcome of a judiciary process in which some use was made of the specific AI tool. Moreover, an attack could be levelled someday against all AI tools applied to the judiciary, on grounds of practical impossibility to ascertain whether some hidden module in some of the tools around contains code thus misconceived... Regardless of counterclaims about error containedness estimates or even proofs, or then about the possibility of making the workings of a given tool even for the uninitiated, I feel it's important for AI researchers to be fully aware of all the implications before they embark on design choices in AI & Law.

2 "Probability" and Narrative

Judging whether the plot of a story is probable or improbable is not necessarily confined to crime stories, even though the next quotation is somewhat relevant to the way my own ALIBI system makes up expalations (or pretexts); James R. Scafidel discussed (1995, p. 358) a short story by Eudora Welty, "Where is the Voice Coming From?". Scafidel was reviewing a book by Gretlund on Welty:

> To the sensitivities of a fellow Mississipian who concocts and publishes mystery novels for a (lowly) living, the plot of [the story] seems highly improbable. Why would this cold-blooded assassin leave his rifle at the scene of the crime? In real life, inexplicable things happen. But this is not real life. If a character in a story does such an odd thing, the author ought to provide some believable reason for his doing it. The narrator's explanation may be cleverly metaphorical, but is still silly: the gun barrel was hot. Surely a man who is clever enough to turn such phrases as "fixes on me like a preacher's when he's yelling 'Are you saved?' " would have enough sense to handle a hot gun by the stock. Gretlund's explanation, that the Voice is leaving a clue so that he will be arrested and "credited" with the crime, seems strained. Why leave a clue? Why not hang around and be caught?.

Kramnick's (1994) is an enlightening essay on eighteenth-century English literary criticism and its attitudes to the plots of novels as a genre, as opposed to the plots of Shakespeare's plays. In her *Shakespeare Illustrated* (1753), Charlotte Lennox provided a translation of thematic precedents (or sources) of Shakespeare's plays; an advocate of the superiority of the novel genre of her own times over Shakespeare's plots, whose plausibility or "probability", she argued, was lesser, she somewhat paradoxically or proleptically (anachronistically) kept referring to Shakespeare's sources as "novels and histories". Kramnick devotes much attention to her sense of the "probability" of the plots (p. 433):[4]

> Lennox's theorization of the novel in *Shakespeare Illustrated* begins by comparing the comedies to their largely Italian sources. Like Fielding in *Tom Jones*, Lennox centers the elaboration and defense of the novel on "probability". In the probable unfolding of the narrative (or "fable", "story", "action", and the like) and the probable delineation of character, Lennox finds the generic standards to evaluate Shakespeare's plays and their sources.

Kramnick also outlines the cultural context in relation to the emergence of early research in statistics (Kramnick, *ibid.*, fn. 10):

[3] Albin Eser of the University of Freiburg concluded a comparatist paper (1997) on the collection and evaluation of the criminal evidence, with a long footnote *(ibid.*, p. 438, fn. 21) using football as a metaphor for American versus European attitudes to the exclusionary rules of evidence: "On the one hand, it is characteristic of European soccer that the players seek out a path to the goal that is as direct as possible and that the flow of the game be held up by as few rules as possible. As a result, spectators are annoyed when referees interrup the game on account of rule violations. In contrast, in American football, the setting up of and compliance with a complicated set of rules seems almost to be a central goal of the game. The impression created by the American criminal trial—with its complicated rules of evidence that lead to endless interruptions—when compared to the Continental-European system, in which the judge is hindered only by comparatively few rules from pursuing an investigation of the truth that is free from interruption, is not much different. William Pizzi was so fascinated by this comparison of European and American systems of criminal procedure and sport, which I voiced in a conversation with him in late 1993, that he has since subjected it to a closer analysis. See [Pizzi (1995)]" *(ibid.*.

[4] In 18th and 19th century English narrative, there actually has been a tradition of structuring the narrative with circumstantial evidence inserted to guide the reader by telling things not seen, making the facts speak for themselves as though (Welsh 1992, Copley 1994).

3 The Doctrine of Chances & Uncharged Conduct

Where does the divide run, in law, between admitting that a coincidence is just that, and assuming causality? What about the admissibility of character evidence—of past, yet uncharged conduct on the part of the accused?[7] In the *New York Times*, George Judson (1995) reported from Owego, N.Y.: the accused, an

> 48-year-old woman accused of smothering her five infant children a quarter of a century ago, was convicted of their deaths today in Tioga County Court. In 1972, a leading medical journal cited the deaths of two infants from rural New York, "MH" and "NH," as compelling evidence that Sudden Infant Death Syndrome ran in families. Today, a jury found that the babies [...] were murdered by their mother, as were two brothers and a sister before them.

The defendant "had confessed to state troopers last year that she had smothered her babies", "a chilling and detailed confession", yet according to her she "had testified that she made the confession only to end hours of questioning by state troopers, saying that her children had simply stopped breathing, sometimes even as she fed them" *(ibid.)*. According to the confession, the babies crying spells were the trigger; in contrast, she "suggested in her confession", the boy she and her husband adopted afterwards is alive as "unlike the five others, he had survived his crying spells because his father was out of work and at home during his infancy, and she had not been left to cope with the child alone". The five murder verdicts are of murder by depraved indifference (i.e., without the conscious intention to kill). The Hoyt case from upstate New York "was striking [...] also for the family's place at the center of research" which at the time was prominent in promoting a medical theory on cot deaths *(ibid.)*. "But to a forensic pathologist in Dallas, [...] the death of five children in one family from SIDS was statistically impossible, and she believed that [the aforementioned] research was leading pediatricians to disregard danger signs within some families" *(ibid.)*. Benderly (1997) approaches the effect of the Hoyt diagnosis of old and recent multiple murder verdicts from the viewpoint of scientific error and its effects on subsequent research.

Again on the Hoyt case: "Criminal defense lawyers know how difficult it is to overcome a confession in a criminal trial, for juries find it hard to fathom why anyone would falsely implicate oneself" (Williams 1996). Yet, generally speaking: "Confessions are usually used as ground truth but are not 100 per cent reliable": so Vrij (1998a: p. 89), who discusses experiments with polygraph examinations. In fact, even "people considered as guilty by virtue of a confession may actually be innocent, as some innocent people do confess" *(ibid.)*.[8]

Prof. Tillers also kindly referred me (p.c., 9 Feb. 2000) to the "famous case 'Brides in the Bath' ":

> Rex v. Smith, 11 Cr. App. R. 229, 84 L.J.K.B. 2153 (1915) (husband perhaps drowned a number of wives to recover insurance proceeds; at first sight the drownings were accidental but...)

Moreover, he sent me an article by a professor of Law from the University of California at Davis, Edward Imwinkelried (1990), a paper which "has an extensive discussion of the American view of the 'doctrine of chances'." Imwinkelried's paper, "The use of evidence of an accused's uncharged misconduct to prove

opposed to ethnic background) in shaping the perpetrators' motives and decision-making behind a hate-crime of a murder of the circumstantial magnitude of setting fire to the victim. This in turn is social commonsense about hate crimes.

[7] On the day I was typing the rest of this section (now prefaced with this paragraph), our School Office sent staff here e-mail about the progress of a colleague who unfortunately was in need of surgery; the message ended by signalling a coincidence: "He did thank all of you who signed his card, amazingly it was painted by his dentist. I found the card in a back street newsagent here in Greenwich!".

Daniel R. Williams (1996), reviewing in the *New York Law Journal* a book written by three journalists on the Hoyt trial from upstate New York on the deaths of five infants, starts by this remark: "Prosecutors resist the idea that coincidences can explain incriminating circumstances, echoing the sentiment of British writer J.B. Priestley who once said that 'although we talk so much about coincidence we do not really believe in it, for in our heart of hearts we think better of the universe'." If it's not coincidence, we may add, we think worse of human nature, when the event is an unfortunate one which we come to believe is a crime committed by some perpetrator. Some other times, the criminal character of an event is not beyond dispute, yet, as in the Jama case, by coincidence the "hidden blessing" obtains (we'll come back to that) by which exceptionally "special" eyewitnesses are available, and this exceptionality affects our expectation—or our wish to to believe—that their testimony will result in punishment for the perpetrators.

I asked Prof. Peter Tillers (from the Cardozo School of Law in New York) to provide me with a reference to cases from the United States which somewhat resemble the recent British case of a mother being found guilty as her second baby also died a cot death (Sudden Infant Death Syndrome, or SIDS). After kindly searching the Lexis-Nexis database, he referred me (p.c., 10 Feb. 2000) to several papers; among these, Williams's review I quoted from above, as well as (on the same trial) Judson (1995) and the other works referenced in this section.

[8] There also is a section on false confessions in Vrij (1998b), from a British coign of vantage. Vrij (2000) is on the psychology of lying and its detection. Cabras (1996) provides a discussion of the interplay of the media, expert witness, and the inquiry in the case of the Monster of Foligno, a serial killer who was targeting children releasing messages from time to time. Eventually the culprit was apprehended; previously, a man had come forth and incriminated himself with a false confession. "Until recently British judges were reluctant to admit expert evidence on the reliability of confessions where the defendant is 'normal', even though his or her personality may be vulnerable" (McEwan 1995: pp. 503–304).

> We may consider Lennox's work, therefore, as part of the process in which, according to Douglas Patey, eighteenth-century critics elevate "probabilistic inference" from its early association with rhetoric and lesser truth to a type of knowledge continuous with scientific, or "demonstrative", certainty. Lennox's particular sense of "probability", further, seems to square with Patey's claim that the discourse of "probability" had, in addition to a central, mediating status in "Augustan literary theory", a particular grounding in the emergent novel, which was understood by the midcentury to combine and condense probabilistic theories of narrative with those of character (Patey [1984], 89, passim). Patey's thesis wants to square the rise of the novel with neo-Baconian accounts of "probabilistic inference"; Lennox's understanding of the term is far more flexible and in fact, demonstrates the resiliency or recrudescence of the Aristotelian model.

There is no room here to overview the early connections of probability theory with, respectively, law and literature. Skipping a few centuries, let it suffice to mention that there is a fierce ongoing debate, among legal evidence scholars, between the so-called "Bayesian enthusiasts" and the "Bayesio-skeptics". On this controversy, see for example Allen & Redmayne (1997), Tillers & Green (1988). In the range of opinions, the most extreme authors in the Bayesian camp, appear to be Robertson and Vignaux; see their book (1995). Almost all authors in the Bayesian camp admit that except in DNA-related evidence, statistics is not to be resorted to in the courtroom, at least openly (in explanations to a jury). Even the Bayesian enthusiasts concede that one thing is using Bayesianism in court, another thing is studying a case in other circumstances, e.g., in retrospect as in Kadane and Schum's (1996) generally acclaimed analysis of the controversial Sacco and Vanzetti case from U.S. early twentieth century history. It took several years to piece together that impressive analysis, not a feasible practice for the courtroom, even though a few argue otherwise, advocating the application of statistics to the dynamics of marshalling the evidence. Anyway, that a correct understanding of the basics of statistics is necessary for the courts can be seen if we are to consider a recent case from the U.K., in which a career woman (a lawyer) was condemned for the deaths of her two babies (she and her husband deny wrongdoing on her part), based on the assumption that the odds of two separate cot deaths occurring in the same household militate against finding her not guilty. (Yet, even disregarding the need to question statistical independence, note that even probability zero does *not* mean an event could not occur.) The *doctrine of chances* has a place in legal cases, and involves admitting or not admitting evidence against the accused[5] from previous *uncharged conduct:* see on this Sec. 3 below. I couldn't tell whether considerations on the admissibility for evidence of uncharged conduct on the part of the individual suspects were explicitly involved in the inquiry on the Jama case.[6]

[5] The burdens of proof are all important. Scientific uncertainty and burdens of proofs in, respectively, scientific practice and environmental law are discussed—from the vantage point of the philosophy of science—in Lemons et al. (1997).

[6] Anyway, the Jama murder case is different from the kind of criminal cases for which the applicability of Bayesian approaches has been debated in the legal evidence literature between the Bayesian enthusiasts and the Bayesio-skeptics: the awkward coincidence in the Jama case is that several football referees and linesmen happened to witness a crime (which took place outside a stadium or playground); and that this awkward coincidence happened is a *given.* This is the point of my argument that this "improbable" element of the narrative makes the latter fall in a pattern of "life imitates art", where art happens to be, in this case, a plot *à la* Brecht. Regardless of the case at hand (I'm no exception to the general public perception that the suspects of the Jama murder should have stood trial), this "intertextual" reference from life to the literary canon can be expected to have triggered reactions agreeing or anyway not ignoring the *telos* (the idea poetically conveyed into the text) of Brecht's plays, of for that matter more generally of texts with a social message.

For the case at hand, the telos is an incentive factor confirming, to the public mind, the miscarriage of justice hypothesis, with a constellation featuring the (to Brecht, class-related) group identity of the victim and the suspects "determining", as though, the outcome by institutional intervention regardless of the (likely) perpetrator status. What focuses the attention on such a pattern is that the what would have appeared to be a formidable procedural obstacle to outcome "release of the suspects"—the testimony of apparently good, even grotesquely good, and for that matter numerous eyewitnesses—was not effective to prevent that outcome, with the inquiring magistrate's public comment "These are good lads". Was that a lemma to the conclusion of the inquiry, or rather a piece of evidence (the character evidence about the suspects)? On the "evidence" of such plays as Brecht's *The Exception and the Rule,* one would be tempted to say the latter applied in the case at hand. I argue that regardless of this being or not being a good inferential practice, in AI models of *agents' beliefs* (Ballim and Wilks 1991)—about legal narratives in particular (cf. Barnden, in press; Ballim et al., 1990 & in press)—including public beliefs and possibly embedded beliefs, this kind of cultural factor would be usefully taken into account. We say more on that in Section "An Input from AI" in Part I of this paper.

Concerning the hate-crime hypothesis for the Jama murder, something that was not made explicit in the JAMA prototype yet calls for statement is that as Jama, the victim, was sleeping on the bench in the park, to observers not acquainted with him (and observers include the perpetrators of his murder, not just the eyewitnesses, who may even not be close enough to notice some details), one characteristic that (we can assume) was conspicuous is that he was a Black man. Another characteristic suggested by the circumstances of the given situation (as well as by contextual social common knowledge) is socio-economical marginality (a homeless immigrant). The victim's biography, educational and professional background, and so forth (Jama had been a lawyer and a teacher before leaving his country of origin) would neither be an association suggested by the sight of the sleeping man, nor be of much relevance (as

mens rea: the doctrines that threaten to engulf the character evidence prohibition", states: "The admissibility of uncharged misconduct evidence is the single most important issue in contemporary criminal evidence law. The issue has figured importantly in several of the most celebrated criminal trials of our time". The introduction starts by describing a hypothetical case in which:

> The accused is charged with homicide. The indictment alleges that the accused committed the murder in early 1990. During the government's case-in-chief at trial, the prosecutor calls a witness. The witness begins describing a killing that the accused supposedly committed in 1989. The defense strenuously objects that the witness's testimony is "nothing more than blatantly inadmissible evidence of the accused's general bad character." However, at sidebar the prosecutor makes an offer of proof that the 1989 killing was perpetrated with "exactly the same modus operandi as the 1990 murder." Given this state of the record, how should the trial judge rule on the defense objection?

Federal Rule of Evidence 404(b)—"which is virtually identical to Military Rule 404(b)" (the paper was published in the American *Military Law Review*— "forbids the judge from admitting the evidence as circumstantial proof of the accused's conduct on the alleged occasion in 1990. [...] Thus, the prosecutor cannot offer the witness's testimony about the 1989 incident to prove the accused's disposition toward murder and, in turn, use the accused's antisocial disposition as evidence that the accused committed the alleged 1990 murder". Yet, the judge is permitted "to admit the evidence when it is relevant on a noncharacter theory", as

> "uncharged misconduct evidence 'may, however, be admissible for other purposes, such as proof of motive, opportunity, intent, preparation, plan, knowledge, identity, or absence of mistake or accident.' In our hypothetical case, the trial judge could allow the prosecutor to introduce the 1989 incident to establish the accused's identity as the perpetrator of the 1990 killing. If the two killings were committed with the identical, unique modus operandi, the uncharged incident is logically relevant to prove the accused's identity as the perpetrator of the charged crime without relying on a forbidden character inference. Hence, the judge could properly admit the testimony with a limiting instruction identifying the permissible and impermissible uses of the evidence.

The text as quoted below is the way it is excerpted in the the summary:

> Unless the judge clearly explains the law governing stipulations, a juror might suspect that any accused who knew enough about the crime to stipulate to the *mens rea* must have been involved personally in the crime. [...] When the question is the existence of the *mens rea*, the prosecutor ordinarily has a much more compelling need to resort to probative uncharged misconduct evidence. [ldots] The character evidence prohibition is violated when we permit a prosecutor to rely on the theory depicted in Figure 2 to justify the admissibility for uncharged misconduct evidence. [...] The courts should admit uncharged misconduct evidence under the doctrine to prove *mens rea* only when the prosecutor can make persuasive showings that each uncharged incident is similar to the charged offense and that the accused has been involved in such incidents more frequently than the typical person. [...]

Imwinkelried's (1990) stated purpose in his paper

> is to describe and critique [...] two lines of authority. The first section of the article discusses one line, namely, the case law advancing the proposition that the first sentence in Rule 404(b) [namely: "Evidence of other crimes, wrongs, or acts is not admissible to prove the character of a person in order to show action in conformity therewith"] is automatically inapplicable whenever the prosecutor offers uncharged misconduct to support an ultimate inference of mental intent rather than physical conduct. The next section of the article analyzes the second line of authority. That line includes the decisions urging that under the doctrine of objective chances, the prosecutor routinely can offer uncharged misconduct on a noncharacter theory to prove intent. Both lines of authority are spurious, and both represent grave threats to the continued viability of the character evidence prohibition.

Prof. Tillers also referred me to the news from the *New York Times* of Sunday, March 19, 1995; in the words of the report—from Hot Sulphur Springs, Colorado—"A woman whose 11 marriages earned her the nickname the Black Widow was convicted on Friday of torturing and killing her ninth husband" *(NYT* 1995). This particular husband had "hired a private investigator when he began to suspect that she was lying about how many times she had been married", and had intended to sue her for fraud and emotional distress. She was divorced from all previous husbands, except the eighth (her marriage to the ninth was annulled for that very reason), and "except for an elderly man who died of natural causes", and this includes her having divorced (twice) from "the lawyer who helped her avoid questioning in the 1972 shooting death of her third husband". In closing arguments, defense lawyers denied there was any physical evidence to dismiss the alibi of the two defendants (the woman and her boyfriend, also convicted,

had claimed they had been away, camping). As to the admissibility of character evidence, it is remarkable that the two defendants were convicted even though "[t]estimony about her previous marriages was not allowed during her trial" *(ibid.)*.

4 The Bayesian Debate in Law

That proof of guilt is, alas, dramatically different from proving a mathematical theorem is nothing new for sure. James Ogden, of the University College of Wales, Aberystwyth—in a note on Othello's demand to Iago, "Give me the ocular proof" *Othello* III.iii.365) being echoed (comically, for that matter) in late seventeenth-century plays—remarks: *"Othello* was one of the most popular plays after the Restoration: some twenty revivals are recorded. Thomas Rymer [(1692)] noted that 'from all the Tragedies acted on our English Stage, *Othello* is said to bear the Bell away', yet to Rymer himself it was 'a Bloody Farce' which 'may be a lesson to Husbands, that before their Jealousie be Tragical, the proofs may be Mathematical'."

The debatants in Allen and Redmayne (1997) are legal scholars and statisticians. Two lead articles represent the contenders: Bayesio-skeptic Ron Allen's, and Bayesian Richard Friedman's. Couldn't it be the case that statistics is being fetishized, and that the prestige the tool is accorded is mesmerizing its way into misapportioning its role for the purposes of the scholarship or even the practice of evidence at trials? If that is the case indeed, not only AI will have to develop integrated approaches to plausibility instead of mere probability, but it must resist the lure of promising more that it could ever deliver.[9] As to suggestions "what might juridical proof be if it is not Bayesian", Allen tries to delineate requirements for "plausible explanations", in terms of "such variables as coherence, consistency, completeness, uniqueness, economy and (yes) probability. What I think is occurring at trials, and should occur, is that parties identify their stories and try to support them with evidence, although not always in that order. [...] In a civil case, [the most plausible] hypothesis wins. In a criminal case, if there is a plausible hypothesis of guilt and none of innocence, the state wins; if there is a plausible hypothesis of innocence, the defendant wins. This suggests that the legal debates should move from Bayesianism to plausibility and explanations, a move perhaps the Bayesians would welcome" (274). Friedman remarks: "There are variations among Bayesioskeptics, as among Bayesians. Some Bayesioskeptics are rather uncompromising: for example, Alex Stein [...] has contended that subjective probability theory is 'vacuous' and lacks even heuristic value. Other Bayesioskeptics [e.g., Callen] acknowledge some limited usefulness for Bayesian methods but spend considerable energy criticising those methods and little or none using them" (276, fn. 1). To Friedman, "in litigated matters [...] uncertainty is of a subjective nature", and "whatever the value of Bayesian methods as opposed to classical statistical methods in scientific inquiry, in litigation I believe that a subjectivist approach to probability is the only one that can offer any hope of assisting in the analysis of juridical proof" (277). Adopting a perspective of seeking greater expected utility, and formulating accordingly the degree of confidence as the standard of persuasion, p and d are the plaintiff and the defendant *qua* the party that receives the judgment; $P(\Pi)$ and $P(\Delta)$ respectively represent the probability that the facts are such that the plaintiff—or, instead, the defendant—is entitled to judgment; "for example, $U(p, \Delta)$ equals the social utility of a judgment for the plaintiff when the truth, if it were known, is such that the defendant should receive judgment" (277–278). "In a criminal case, the long-standing and solidly established view is that $U(p, \Delta)$, the negative utility of an incorrect judgment for the prosecution, far exceeds any of the other utilities in magnitude" (278). In civil cases, instead, $U(p, \Pi) = U(d, \Delta)$. Anyway, finding for the plaintiff will be optimal only if the fact-finder's degree of confidence is at least as great as

$$O(\Pi) = \frac{P(\Pi)}{1 - P(\Pi)} = \frac{U(d, \Delta) - U(p, \Delta)}{U(P, \Pi) - U(d, \Pi)}$$

Friedman admits that "the charge is sometimes made of 'Bayesian imperialism'. To this charge, I think, Bayesians should plead half-guilty" (278). He does not believe in the usefulness of alternative systems of probability, leading a fact-finder to results inconsistent with Bayesianism's. However, "just as it is generally best that the audience not see what is going on backstage", Friedman does "not believe that probability theory usually needs to be mentioned in the courtroom" (291). and that, in any case, it

[9] In the words of Allen's article, where the context is about Bayesianism in law: "The landscape of the law has an interesting Darwinian quality. It is littered with dead algorithms and formalisms (henceforth 'algorithms'), but everywhere now algorithms are arising to take their place, and in some instances as with species hastening the extinction. Some of the defunct algorithms and their demise were quite spectacular. Legal science, the idea that the law like a science could be reduced to a set of necessary and sufficient conditions, is one good example [...]. Justice Owen Roberts's theory of constitutional law that to decide the constitutionality of the statute one merely needed to lay the law and the constitution side by side and compare them is another [...]. The [U.S.] Supreme Court's struggle with the varying levels of scrutiny under the equal protection clause is yet another. Less grandly, every field of law of which I am aware is littered with numerous, largely defunct, 'X-part tests' articulated to resolve some general category of problem only to have the tests disintegrate under the pressure of unanticipated developments. The field of evidence is no exception".

needs application to be careful: "If the limited role of Bayesian analysis is kept in mind, then I believe that the arguments of the Bayesioskeptics lose most of their force" *(ibid.)*.

Johan Bring, in his commentary, retorts: "if the limited role of Bayesianism as described by Friedman is kept in mind, the Bayesians' claim loses most of its force!" (292). Bayesianism, to say it with the title of one of the sections in Friedman's lead paper, is "[a]n only partially imperialistic view", and "an incomplete determinant of the standard of persuasion" (279). Whereas he does "not believe that the standard of persuasion necessarily can be expressed solely in terms of probability", nevertheless, to him, it "should play a large role in defining that standard" *(ibid.)*.[10]

Friedman tries to reconcile the story-telling model of fact-finders' tending to view an entire body of evidence, to a probabilistic representation. He had previously remarked (284, fn. 18) that "courts should not be mesmerized by an attempts to determine 'the single most probable story' ", as "there is no clear boundary line to a story; to make the fact-finding process coherent, a wealth of infinitesimally different story lines must be batched together" *(ibid.)*. Computational complexity beyond the powers of human intellect is one more objection to Bayesianism's applicability to legal evidence—"a point that has been especially emphasised by Callen" (288, fn. 26) in several publications—and Friedman claims that "the argument is wide of the mark" (288): "a flexible template", Bayesianism "can take into account as much complexity as its user is able to handle" *(ibid.)*. Athletes cope with a ball hustling through the air without being physicists (289), and likewise, to Friedman, when they think well, fact-finders "reach results that are roughly consistent" with rigorous use of Bayesianism, which anyway (and it may smack of a retreat into the ivory tower) is useful "as an analytical tool" (290).

These are fairly representative viewpoints in the debate. Bring, who criticized Friedman's "story-telling Bayesian model" (292), stated he "do[es] not believe that all scholars have the same sober view of the usefulness of Bayes' theorem as Allen and Friedman" (292). To have a sober say either way, AI researchers will need to learn more. The present readers are warmly encouraged to acquaint themselves with the paper collection itself.[11]

[10] Friedman's addresses Allen's and others' concern with the so-called "problem of conjunction": in Friedman's own terms, "a plaintiff must prove several factual propositions" (279), and difficulties arise when the probability of their conjunction is calculated; moreover, "the number of elements into which a given claim is divided is essentially arbitrary" (280), and "the more elements a claim is divided into, the easier it is for the plaintiff to satisfy the burden with respect to each element" *(ibid.)*. Friedman dismisses the problem by advocating a "cumulative approach", by which "the fact-finder should find for the plaintiff only if $P(A \ \& \ B) > 0.5$" *(ibid.)*; conceding however that "[w]here the defendant raises an affirmative defence, a more complex instruction would be required" *(ibid.,* fn. 9). Then, Friedman sets to explain away Allen's previously published objections to the cumulative approach, and, offering a few more objections himself, proposes to overcome them by policy-based devices (284). The standard of persuasion apart, objections concerning Bayesianism's adequacy to cope with the presumption of innocence are addressed, with Friedman arguing that obviating is feasible, for all of its being "trickier than has been acknowledged by good Bayesians" (285).

Next, Friedman turns to confront the legal narrative aspect *vis-à-vis* Bayesianism, "a false dichotomy" (286). He attributes "[t]he perception of a conflict [...] to a misconception aboput the role of Bayes' theorem in Bayesian reasoning. The theorem is important, but it is not all of Bayesianism" *(ibid.)*. He denies "that fact-finders can, should or do go through [...] a serial updating of probability, given each new piece of evidence" (287), and besides, "not all pieces of evidence call for application of Bayes's theorem" *(ibid.)*.

[11] Friedman's riposte noted: "several leading Bayesioskeptics (Allen, Callen, Stein) acknowledge—with varying degrees of specificity and varying degrees of grudgingness—that standard probability theory can be useful as an analytical tool [...]" (348). "I do not contend that fact finders in litigation ordinarily ought to be required or encouraged to make explicit or conscious use of Bayesian methods at trial. Even Robertson and Vignaux, perhaps the most unyielding Bayesian of all in the evidentiary debate, make clear in this symposium that they agree" (352). Mississippi College's Craig R. Callen, in the comment next to Bring's, "doubt[s] that any of us skeptics would argue that formal analysis is *useless*", yet "we might argue that it chiefly serves to dramatise or illustrate some implications of ideas, which may themselves have no particular formal credentials" (296). Callen's contribution, titled "Computation and juridical proof", "adapt[s] an example which John Searle used to illustrate the difference between minds and computational programs", in order "[t]o show how Bayesianism fails to map many aspects of fact finding in the empirical word" (296). "Bayesianism is a formal computational system", and "such a system cannot accurately represent the complexity of human decision making, in principle or otherwise, without considerable assistance from some other quarter, as yet unidentified by Bayesians" *(ibid.)*. Obviouly, AI research need pay close attention to such an argument, as AI, not just Bayesianism, is within the scope of Callen's critique. "Friedman talks about assessment of the *likelihood* of stories, but not about their formation or development" (297). "Bayesian analysis has nothing to say about whether particular phenomena count as evidence for a particular conclusion. Such decisions, instead, must rely on our own associative memories and interpretive processes, including story formation, which Bayesian analysis does not purport to model" *(ibid.)*. Moreover, "there is no Bayesian process for ascertaining the probative value of a particular piece of evidence, but merely a check to make sure that whatever subjective probabilities one uses are consistent" *(ibid.)*. Also, the inferential process itself is not properly addressed *(ibid.)*. I in turn propose that there is room here for intelligent technologies, including such that even Searle would perhaps not be displeased with. After all, AI and subsymbolic computation have something to offer on those very counts.

References

Allen, R. and Redmayne, M., eds. (1997) *Bayesianism and Juridical Proof.* Special issue of *The International Journal of Evidence and Proof.* 1. London: Blackstone. [Vol. 1 includes 4 regular, numbered issues and one thematic issue, unnumbered.]

Ballim, A. and Wilks, Y. (1991) *Artificial Believers: The Ascription of Belief.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Ballim, A., Wilks, Y. and Barnden, J. (1990) "Belief ascription, metaphor, and intensional identification". Ch. 4 in: S.L. Tsohazidis, *Meanings and Prototypes: Studies in Linguistic Categorization.* London: Routledge, pp. 91–131.

Ballim, A., By, T., Wilks, Y. and Liske, C. (in press) "Modelling agent attitudes in legal reasoning". In Martino and Nissan (to appear).

Barnden, J.A. (in press) "Uncertain reasoning about agents' beliefs and reasoning". In Martino and Nissan (to appear).

Benderly, B.L. (1997) "Turning a blind eye to mad science". (Review of: R. Firstman and J. Talan, *The Death of Innocents,* Bantam.) *The Washington Post,* November 17, 1997, final edn., Section "Style", p. C08.

Bluestone, H., Travin, S. and Marlowe, D., eds. (1994) *Psychiatric-Legal Decision Making by the Mental Health Practitioner: The Clinician as de facto Magistrate.* New York: Wiley.

Bond, C. (1994) "Representing reality: strategies in realism in the early English novel". *Eighteenth-Century Fiction* 62(2) [McMaster Univ., Hamilton, Ontario]: pp. 121–140.

Cabras, C., ed. (1996), "Un mostro di carta". [On the Mostro di Foligno.] Ch. 11 in C. Cabras (ed.), *Psicologia della prova.* Milan: Giuffrè.

Calistri, R.J. 1990. "Classifying and Detecting Plan-Based Misconceptions for Robust Plan Recognition". Ph.D. Dissertation (Computer Science). Technical Report No. CS-90-11. Providence, RI: Brown University.

Copley, S. (1994) Review of Welsh (1992). *Textual Practice* 8(3): p. 526.

Eser, A. (1997) "Collection and evaluation of evidence in comparative perspective". *Israel law Review* 31(1–3): pp. 429–438.

Imwinkelried, E.J. (1990) "The use of evidence of an accused's uncharged misconduct to prove *mens rea:* the doctrines that threaten to engulf the character evidence prohibition". *Military Law Review* [Washington, D.C.: Headquarters, Dept. of the Army, Supt. of Docs.], vol. 130 (Fall 1990): pp. 41–76.

Judson, G. (1995) "Mother Guilty in the Killings of 5 Babies". *The New York Times,* April 22 (Late Edn., Final): Sec. 1, p. 25, col. 5, Metropolitan Desk, Second Front.

Kadane, J.B. and Schum, D.A. (1996) *A Probabilistic Analysis of the Sacco and Vanzetti Evidence.* New York and Chichester: Wiley.

Kramnick, J.B. (1994) "Reading Shakespeare's novels: literary history and cultural politics in the Lennox-Johnson debate". *Modern Language Quarterly* 55(4), pp. 429–453.

Lemons, J. Shrader-Frechette, K. and Cranor, C. (1997) "The precautionary principle: scientific uncertainty and Type I and Type II errors". In: M. Kaiser (ed.), *The Precautionary Principle and Its Implications for Science.* Special issue of: *Foundations of Science* 2(2): pp. 207–236.

Lonsdorf, R.G. (1995) Review of Bluestone et al. (1994). *The Journal of Legal medicine* 16(2): pp. 319–324.

McEwan, J. (1995) "Adversarial and inquisitorial proceedings", Ch. 6.1 in: R. Bull and D. Carson (eds.), *Handbook of Psychology in Legal Contexts.* Chichester: Wiley, pp. 495–508.

Memon, A., Vrij, A. Bull, R. (1998) *Psychology and Law: Truthfulness, Accuracy and Credibility.* London: McGraw-Hill.

Morse, D. (1987) *American Romanticism.* Vol. 1: *From Cooper to Hawthorne: Excessive America.* Vol. 2: *From Melville to James: The Enduring Excessive.* (Macmillan Studies in Romanticism.) London: Macmillan.

Nissan, E. (1987) "Exception-admissibility and typicality in proto-representations". In: H. Czap and C. Galinski (eds.), *Terminology and Knowledge Engineering (Proceedings of the First International Conference,* Trier, 1987). Frankfurt/M: Indeks Verlag, pp. 235–267.

Nissan, E. (1995) "Meanings, expression, and prototypes". *Pragmatics and Cognition* 3(2): pp. 317–364.

Nissan, E. (1996) "From ALIBI to COLUMBUS". In: J. Hulstijn, A. Nijholt (eds.), *Proceedings of the 12th Twente Workshop on Language Technology (TWLT 12).* University of Twente, Netherlands, Sept. 1996, pp. 69–85.

NYT (1995) "Woman guilty of murdering husband no. 9". *The New York Times,* March 19 (Late Edn., Final): Sec. 1, p. 31, col. 1, National Desk.

Ogden, J. (1992) "Restoration jocularity at Othello's expense". *Notes and Queries* [Oxford University Press], year 237, new series, 39(4): p. 464.

Patey, D. (1984) *Probability and Literary Form: Philosophic Theory and Literary Practice in the Augustan Age.* Cambridge: Cambridge University Press.

Pizzi, W.T. (1995) "Soccer, football and trial systems". *Columbia Journal of European Law* 1: pp. 369–377.

Robertson, B. and Vignaux, G.A. (1995) *Interpreting Evidence: Evaluating Forensic Science in the Courtroom.* Chichester: Wiley.

Rosoni, I. (1995) *Quae singula non prosunt collecta iuvant: La teoria della prova indiziaria nell'eta' medievale e moderna.* (Università di Macerata, Pubblicazioni della Facoltà di Giurisprudenza, Seconda serie, vol. 84.) Milan: Giuffrè.

Rymer, T. (1692) *A Short View of Tragedy,* in: C.A. Zimansky (ed.), *The Critical Works of Thomas Rymer.* Yale University Press, 1956.

Scafidel, J.R. (1995) "Eudora Welty's place in time". *Mississippi Quarterly* 48(2): pp. 351–360. [Mississippi State, MS: Mississippi State University.]

Vrij, A. (1998a) "Physiological parameters and credibility: the polygraph". Ch. 4 in Memon et al. (1998).

Vrij, A. (1998b) "Interviewing suspects". Ch. 6 in Memon et al. (1998).

Vrij, A. (2000) *Detecting Lies and Deceit: The Psychology of Lying and Implications for Professional Practice.* (Wiley Series on the Psychology of Crime, Policing and Law.) Chichester: Wiley.

Welsh, A. (1992) *Strong Representations: Narrative and Circumstantial Evidence in England.* Baltimore: The Johns Hopkins University Press.

Williams, D.R. (1996) "Goodbye, My Little Ones" (book review). *New York Law Journal,* April 30, Section "The Lawyer's Bookshelf", p. 2.

RULES AND PRINCIPLES IN LEGAL REASONING. A STUDY OF VAGUENESS AND COLLISIONS IN ARTIFICIAL INTELLIGENCE AND LAW

Samuel Meira Brasil Jr
FDV - Faculdades Integradas de Vitória
Espírito Santo State - Brazil
samuel.brasil@pobox.com

Abstract

There is a difference between rules and principles in Law Argumentation Theory. The conflict of rules is solved by meta-rules like *Lex Superior* (based in general hierarchical structure of legal system), *Lex Posterior* (based on later rule priority), *Lex Specialis* (based on specification of rules) and also by exceptions, which exclude the conflict. The collisions of rules imply the validity negation of one rule. However, this does not occur in a collision of principles. Even when one principle is not applied by collisions, its validity remains. Another point is that the priority could change, changing the overridden principle. So, it is necessary to build a reasoning model that allows reasoning with conflicting legal principles. Another problem is vagueness of principles. Vagueness does not mean just an imprecise meaning, linguistically talking. There should be attributed a semantic, or a pragmatic meaning. Vagueness of principles occur in account of its opened antecedent, with no established conduct. In this paper a model of reasoning using AI&Law methods is described aiming of treating vagueness and collisions of principles, searching to establish rationality in legal argumentation systems.

1 Introduction

Legal norms are formed by rules and principles. Principles, just like rules, establish deontic propositions. But principles, in legal domain, only had a close attention in earlier discussion.

Rules establish a conditional program, with an antecedent which describes an action (condition), and a consequent which prescribes a legal consequence (conclusion). Principles have a similar logical structure, but their program is intended for a purpose. Thus, principles directly establish purposes which the rules intend to achieve.

2 Rules and Principles

The first reference about principles was by Jean Boulanger (1949, 1950). However, it was Esser (1964) who defined and inserted principles concepts in the context of legal discourse.

In Law Theory, Dworkin (1978) had a fundamental importance in distinguishing principles from rules. He claims that differences are mainly caused by the fact that unlike principles, rules have absolute obligations equivalent to "all-or-nothing". Principles do not have this kind of absolute obligations, and they could be used in a future case, even when they are overridden. To Dworkin, the selection or hierarchy of principles is established in a dimension of weight, or in the importance of value. But Dworkin does not accept that a rule prevails over another rule, by dimension of weight criterion, just like principles do.

Alexy (1985) agrees with Dworkin's ideas of principles, but he offers his own approach, in certain relevant points. A mode of distinguishing rules from principles is the abstraction degree on their prescriptions. However, Alexy claims that the distinction is not only a matter of degree, but also a matter of quality. Thus, a conflict of rules is solved in validity domain, while collision of principles is solved on dimension of value.

Peczenik (1971) shows as distinction criterion, the importance of principles to legal system. He assumes that *"principles are normative propositions. They are not descriptive statements. They are what ought to be and what is permitted, not what actually is the case".*

Verheij (1996) has an integrated view on rules and principles. According to Verheij, the difference between rules and principles are merely gradual, since both rules and principles generate reasons. He recognises no

difference in logic structure between rules and principles, just like Soeteman (1991). The system foundation of his thought is Reason Based Logic (RBL), for the representation of rules with exceptions and weighing of reasons (Hage, 1991; Hage and Verheij, 1994).

In my opinion, there are some aspects about principles and rules, which should be pointed, to the scope of this research. First, principles do not directly establish an action on its antecedent, unlike rules. Principles prescribe goals which ought to be achieved. So, principles establish goals, which could be performed by an action, making possible the legal consequence (conclusion). In summary: principles intend to realise aimed situations. Second, a principle prevails over another, on account of the weight of legal value in dispute, which brings the priority problem. Third, principles and rules are prescriptive propositions. The *modus* with which a concept is constructed can result in a principle or in a rule.

But two points have fundamental importance to the analysis of rules and principles. They are vagueness degree, in account of open texture of legal concepts, and behaviour when there occurs a collision.

2.1 Open texture and Vagueness

It is clear that a legislator can not foresee the entire future, to classify all possible situations which may occur in real life. And even when a definition is given, it does not mean that a description of the actual actions is complete. Legal rules are open texture. It is well-known Hart's (1958) distinction between 'core cases' ('clear cases') and 'penumbra cases' ('hard cases'), which is based on the availability of legal knowledge. In clear cases, it is possible to decide if the current case is an instance of the legal concept, based on the information of a statute or judicial precedents. In hard cases, the judicial decision does not have additional knowledge to classify a fact situation. So, in hard cases, whether a certain case is an instance of a legal concept this is uncertain. This occurs on account of the open texture of the legal norm.

It could be observed that open texture is not a particular characteristic of principles. It occurs also for rules. Hart (1958), *apud* Prakken (1997), cites as an example a park regulation, forbidding the use of vehicles in the park. Although been a rule, a problem in classifying 'vehicles' exists, because its definition is not complete. It could be a skateboard, a car or anything else. So, rules also could have an open texture.

What could be attributed as a relevant criterion of distinction is the vague degree of the action described in the antecedent of the rule or principle.

In rules, the antecedent establishes a description of an action, although in a not complete definition. The action described could be open texture. But although an abstraction degree exists, an action is described in the antecedent. So, the antecedent is determined, though uncertain, and sometimes open texture. In this case, the antecedent should be interpreted, viz. a rule that allows a divorce, when been marriage becomes insupportable. We could ask: what is the definition of 'insupportable'? What actions could be classified as 'insupportable'?

In principles there is not an action described in the antecedent. And this, because in principles, the actions are the conditions for the achievement of a goal, as will be discussed. Thus, any sufficient action to realise the purpose of a principle could be classified as antecedent of that principle. The antecedent of a principle is completely open textured, or uncertain, and should be constructed.

The abstraction degree of a principle when compared with a rule, increases considerably.

An important note done by Prakken (1997), is what he calls, as Gordon (1991), 'underdetermination' and 'overdetermination' of open texture legal rules. According to Prakken, 'underdetermination' is a kind of open texture, concerning a problem of content:

A problem of whether a certain fact situation can be classified as an instance of a legal concept on the basis of established legal knowledge.

However, if conflicting classification rules exist, the open texture is 'overdetermination':

...there is a related problem of open texture which indeed causes problems for logic. It is likely that when legal doctrine develops on the issue, after a period of time a body of conflicting information exists, consisting of judiciary decisions in the first cases, expert opinions, dictionary interpretations, etcetera.

With 'overdetermination' there is inconsistent information.

Besides these two kinds of open texture, I shall claim a different one. Even when a concept has its semantic meaning clearly classified, conceptually talking, it remains open texture. The society is dynamic, on its cultural values. And some behaviours foreseen on legal system change over the years. These modifications of behaviours by culture could imply a legal text modification, by legislator. But such modifications could happen without any textual modification. They could happen with the modification of the Judge interpretation, modifying the concept previously adopted, even when the legal text remains the same. In Constitutional Law, we should call such phenomenon as Constitutional Mutation (Verfassungswandlung).

Thus, even when there is no textual modification, there could be conceptual modification. Or, in other words, if a case could be classified as one of that kind previously decided, the mutation of the meaning could exclude a conclusion. After Prakken (1997) and Gordon (1991), I shall call such open texture as *alterdetermination* .

So, even a clear case remains open texture by modifying the conceptual interpretation, following the social mutation over the content of the concept.

This possible mutation does not need to be implemented in all legal argumentation, or at least in all task of a system. Perelman (1958) claims an inertia principle, which just allows the modification of the decision, if new and sufficient reasons are presented for that.

Such approach could be used in legal system. So, the task will only be implemented, if a new situation claims for it.

But the problem concerning open texture concepts remains. Without intending to offer a final solution to this issue, I shall propose an approach based on ideas of Habermas's (1973, 1973a) Truth Consensual Theory. Accordingly, a proposition is true if the validity pretension of an enunciation is justified. The justification of an assertion does not depend on the truth of the assertion, but the truth of an assertion depends on its justification. The concept of truth is transferred from the semantic level to the pragmatic level.

The basic idea of Habermas's theory is that a fact is the content of a proposition which could be discursively justified.

According to Habermas (1973, 1973a), there is no field in discourse to obtain information. Discourses are free of actions and experiences. In discourses, information is only introduced, and its result arises from accepting or rejecting its validity pretension.

Thus, an open texture concept could be classified upon justifying its content.

In essence, this approach does not prevent the open texture from legal rules or principles, but it is a way of selecting a content. On advice, it is important to observe that this approach can not avoid arguments against the content, with exceptions and defeasible reasonings.

2.2 Conflicts and Collisions

Rules may have contradictory conclusions. Once everything could be derived from contradictory conclusions, it is necessary to find a solution for the conflict of rules. The conflict of rules implies the invalidity of one conflicting rule. A criterion to solve this conflict is found in meta-rules like *Lex Superior* (based in general hierarchical structure of legal system), *Lex Specialis* (based on specification of rule) and *Lex Posterior* (based on later rule vigency). The latter is considered, by some legal expertise, a weak criterion. There is another way to solve conflict of rules, which is to include an exception clause. This exception clause modifies the meaning of one of the conflicting rules, through interpretation. The interpretation, changing the semantic content of the concept, avoids conflicting meaning and allows harmony between rules.

In this sense, the conflict of rules implies the invalidity of one rule, or the invalidity of the meaning which is avoided by the exception clause. The exception clause does not invalidate the rule, because it avoids the invalid meaning and, in such a way, it avoids the conflict itself.

As it happens with rules, principles could have contradictory conclusions.

However, it is in the collision of principles which resides the true problem of legal argumentation. The precedence of one principle over another does not mean that the last principle is invalid, as it occurs with rules. E. g., the precedence of the principle of consumer protection over the principle of contract freedom, in a specific case, does not arise the invalidity of the excluded principle.

Even though overridden in one case, a principle could be used in another case with precedence.

An excluded rule, however, could not be used in another case, because its validity is denied. So, a rule does not prevail over another in account of its weight (of reason or goal), unlike principles do. This is mainly caused because such conflict is treated by correcting the validity aspect of the rule.

Thus, conflict of rules are solved on a validity level, while collision of principles are solved on a dimension of value.

So, it is interesting to note that in the collision of principles there is no invalidity conclusion. The principle only will not be used in that case [1].

Furthermore, from this point of view the inconsistency of the system will not arise, in account of two valid and contradictory propositions. In case of a well defined priority criterion, the inconsistency could be avoided.

In a matter of fact, if only one principle is obtained by precedence over another, there will not be inconsistent, because just one conclusion will be possible.

On the precedence of a principle over another principle, or of a principle over a rule, arise the problem about priority reasoning.

2.2.1 The problem of priority

A collision of principles results in pondering rights or legal values which are in dispute. This reflections occurs over the weight conferred to legal value in dispute. In BverfGE 7, 377, the decision of the Federal Constitutional Court was based on a careful reflection over importance of opposite interests.

However, weight of rights is measurable, but it results from values.

Hubmann (1977), *apud* Larenz (1983), tried to make a mathematical calculation about the dimension of weight, but he admits that his method only could be applied with previously numeric evaluation, and this evaluation would still remain arbitrary.

Intuitively, we can find the right weights. However, for the rationality of the discourse, the weight attributed ought to be justified.

It was proposed as a solution to this problem, the Utilitarianism principle. Classic Utilitarianism recognizes just one decisive criterion, based on the utilitarian principle. The basic idea [2] consists in applying to the greatest possible number of people in a society, the principle of rational choice used by an individual.

Rawls (1971) claims an alternative approach over Utilitarianism principle, based on intuitions. He argues that there is no irrationality in invoking intuition to find a solution to priority problem. He recognizes the difficulty of finding an objective way to eliminate the plurality of conflicting principles, or even of the weighing problem.

However, Rawls understands the weighing attribution as essencial, because without weights, there could not be rational discourse. He shows, as a matter of fact, that his thoughts on intuition are only partial. In a word: principles ought to be explicitly formulated to solve priority problem, even when does not eliminate intuition.

Rawls does an interestingly evaluation of the priority problem, and proposes a serial or lexical order, in case of maximal obligatory principles. A serial order consists in weighing the importance of a principle by attributing a serial option. If an anterior principle has not a well defined application, the posterior will not be applied. But such approach does not allow a modification of the priority, by changing the precedence of principles in collision, as claimed in this study.

In my opinion, priority should be on the weighing of the purpose. Principles aim at the achievement of goals or purposes. Goals are desirable situations, but they have deontologic nature.

The implication of a consequent from an antecedent, in the hypothesis of principles, occurs necessarily through the goal.

As the antecedent is open textured, the derivation of the consequent will occur only if the antecedent achieves the purpose, or the goal of the principle.

---

[1] However, there is a possibility of invalidating a principle, when the goal aimed is forbidden by law or social values.

[2] It is not in the scope of this study to examine the evolution on the Utilitarianism concept, but just to present a basic idea of its contents.

Thus, the problem of priority should be solved by weighing the purposes, and not weighing the principles themselves.

In essence, the goals to be weighted are those which a particular action aims at realises, and those which is actually achieved. So, even if there are not subjective intentions to achieve a goal, if this goal is achieved, there is a collisions of principles.

In a word: if the weighing of a purpose of principle 'A' takes precedence in a particular case, principle 'B' is overridden. But it does not mean that overridden principle 'B' is invalid, or that its degree of priority is lower. On the contrary, the overridden principle 'B' could be used in any other case and, even if its goals change, it can take precedence over principle 'A'.

Note the example of an advertisement enhancing a different treatment based on the race or the sex of individuals. It could be argued that it is contrary to the principle of equality, and that it may be even enhancing racial or sexual preconceptions. A counterargument could be based on the principle of freedom of speech. Which principle takes precedence over the other? The principle of freedom of speech? Or the principle of equality?

The principle of equality should prevail, if the *purpose* of the action is the racial or sexual preconception. However, this conclusion does not convey the idea that the principle of equality will *always* have priority over the principle of freedom of speech. If the *goal* changes, the priority will change too. For instance, an advertisement whose purpose is not to show racial or sexual preconception, but is oriented to reach specific ethnical or sexual group, viz. selling something to this ethnical or sexual group. We could not assume that an advertisement of an oriental or na African doll intends to promote preconception. In this case, the goal has changed, and the principle of freedom of speech takes precedence over the principle of equality.

Thus, priority can occur only in relation to goals or the purposes of the principles in collisions, through the actions which achieves them, and not over the principles themselves.

This proposed priority reasoning could be used in the collisions of principles, and in the collisions of principles and rules. A conflict of rules, does not accept the precedence of one rule over the other, based on the weighing of purposes, without any reference to a principle. If this reference exists, the collisions will be between principles and rules.

This approach allows the modification of the priority between principles, which other approaches do not. So, it does not mean a return to the Utilitarianism principle, because the evaluation of final principles are done in different ways [3]. Thus, the difference between individuals are considerable. In the same way, this approach does not mean an appeal to one's intuition, because it uses a weighing criterion to treat the priority problem, bringing rationality to the legal argumentation.

It is, in essence, a teleological reasoning. Rawls see Utilitarianism as a Teleological Theory. He also sees Intuition as a Teleological Theory. But this issue is not biconditional, and the Teleological Theory is not limited to the Utilitarianism or to Intuition.

In Lüth Case (BVerfGE 7, 198), the Federal Constitutional Court [4] decided that the reflection over the rights in dispute should comprehend, in first place, its reasons and goals; in second place, the action should not overstate the necessary and supportable damage. This decision was based on the fact that the purpose to be achieved was not contrary to moral, and the selected way to achieve it would not overstate what was licit.

Although explicit comments about conflicting principles, as proposed in this study, had not been made by the Court, the precedent set in this case showed that the Judges had been convinced by the purposes and the goals in dispute. The Court decided on the priority of legal assets, which were the goals and purposes in collisions.

The decision made in Lüth Case brings up the issue involving the proportionality of legal assets. This matter is also very important in Law Argumentation Theory, but it is not the scope of this study.

In BVerfGE 30, 173 (Mephisto Case of Klaus Mann), the Judge Rupp-v. Brünneck based her decision on the

[3] The evaluation of principles in different ways, usually frequent due to social differences → different justice conceptions, whether to individuals, whether to social groups, what is not allowed by Utilitarianism.

[4] Other decisions made on purposes or goals are: BVerfGE 7, 377; BVerfGE 7, 198; BVerfGE 21, 239 (press freedom against Criminal Law); BVerfGE 27, 71.

purpose of the action, by pondering whether these was intent injury.

2.2.2 Teleological Reasoning

Teleological reasoning, in a simple definition, occurs when a purpose is obligatory, and some actions ought to be performed for the achievement of that goal. So, the actions is justified when the purpose is achieved. Once the achievement of the purpose is the main role of teleological reasoning, the actions that deny that purpose are to be avoided (Hage, 1995).

In summary, principles are legal norms which directly prescribe goals and purposes. Thus, principles do not directly describe an action to be done, but instead, a purpose that ought to be achieved by an action. The action is the way to achieve the purpose.

Purpose is not the same as reason. Reason is the *leit motiv* of the conclusion. After Hage (1995) I would define reason as a fact that has a particular significance for a conclusion. Purpose is the obligatory state aimed at by regulations, or the facts that conclusions should achieve. Reason could not be the purpose of a principle.

Alexy (1978) shows in different formalizations the logical structure of genetic interpretation and teleological interpretation. On the first one, the interpretation is done by the legislator's intention, and what the legislator intends to achieve. He uses this formalization:

For this, it is necessary an inference rule which shows that the interpretation of rule R through W, desired by the legislator, is a reason for R' validity.

· (1) R' (=I↯) is desired by legislator
(2) R'

In teleological reasoning, the purpose is obligatory, and not only desired by the legislator. This is Alexy's formalization of the teleological interpretation:

· (1) OZ
· (2) ¬ R' → ¬ Z
(3) R'

In other word, if Z is obligatory, whatever is necessary to achieve Z is obligatory.

It is important to note that purpose, in teleological reasoning, is an obligatory condition. So, classifying the content of a concept as an issue of a principle, should be done considering the purpose of that principle, as discussed before (2.1). And the priority problem should consider the purpose of the actions in the collisions of principles.

3 Argumentation level

Toulmin (1958) includes in his argument structure 'rebutall', in which defeasible of an argument is allowed, and 'backing', in which a justification is given. Principles or rules could be used as 'rebutall', and could conflict with 'warrant' or even with 'backing'. On the other hand, the 'backing' of an argument could also be a principle or a rule. That is the reason for a priority framework. But this does not prevent legal argumentation possibilities.

The precedence of a principle does not prevent exceptions. An argument could be defeated by stronger counterarguments, even after priority has been achieved. This issue brings up nonmonotonicity. Thus, it is necessary to compare arguments, once they are in conflict. Conflicts can occur (Pollock, 1987; Prakken, 1997) when arguments have contradictory conclusions (rebutting), or when arguments deny an assumption of the other or the link (undercutting). Reasoning can be done, as claimed by other researchers, formalizing nonmonotonic reasoning to construct and compare arguments. In conclusion, a priority framework should be integrated with an argumentation framework.

4 Brief notes on AI & Law systems

I should make a few general remarks on the others reasoning systems, just to show some relevant aspects of this study.

4.1 Case Based Reasoning

Case Based Reasoning (CBR) is useful to compare cases, and could be used in some tasks of the reasoning model being presented. However, CBR is far from finishing discussion about legal reasoning. Legal arguments could be defeated by counterarguments, as shown by Hart (1958) and Toulmin (1958), in legal domain, and by Prakken (1993) in AI&Law, among others. In one word: legal rules, or even legal cases, are defeasible.

Another point to be observed, in the approach of this study, is the possibility of overriding similar cases. A case previously decided, in its all relevant circumstances, could be decided in a different way, in account of the value modification of its relevant consequences. (Alexy, 1978). These issue has been described here as alterdetermination of open texture legal concepts, and can be found in Constitutional Law principle called Constitutional Mutation.

4.2 Argumentation Systems

Argumentation Systems formalizes nonmonotonic reasoning to construct and compare arguments, for alternative conclusions. The defeasibility of an argument by stronger counterarguments from nonmonotonic logic, is the basic idea of such systems.

The approach suggested in this study is not in conflict with Argument-based reasoning. By contrast, it is integrated with it. Weight of purposes is open texture, because its purpose is not explicitly determined. Even when its definition is incomplete, it could be represented in propositional logic. And even a principle with justified purposes,[5] can be defeasible by counterarguments. The Argumentation System developed by Prakken (1997) is very interesting and attractive. And reasoning with purposes has an argumentation level, which could not be ignored. In essence, this approach of purpose reasoning has an integrated view with the Argumentation Systems.

4.3 Reason-Based Logic

Jaap Hage (1996; 1997) and Bart Verheij (1996) described versions of Reason-Based Logic, with which rules and principles are syntactically indistinguishable. In this interesting system, the output is generated by weighing reasons for or against a particular conclusion. RBL includes principles and goals on its primary level, and rules on its secondary level. Accordingly, a rule generates a reason for its conclusion, and could generate a reason against applying the principles that underlie that rule. So, rules could replace the reasons on which they are based on. In summary, Hage's and Verheij's approaches are very interesting logical models for reasoning with weights.
This study is also based on a weighing process, but just on weighing purposes, and not on weighing reasons. I shall distinguish reasons from goals, which allows a different logic treatment. In the same way, in my system, principles have a different approach to rules, on which logical treatment of conflicts and collisions are based on. This study just examines a different approach to principles and rules, its open texture and the priority problem, which is integrated with argument-based reasoning, despite the attractiveness of such a system like RBL.

Like RBL, the knowledge of reasoning with goals is stated as logical inference rules, and not as premises. It is important to note that Hage (1992) has already researches on teleological reasoning, but he restricts his researches to RBL approach.

4.4 Prioritized Default Logic

Brewka (1994a) developed a prioritized default logic, on which he includes reasoning about priorities. His system is based on priorities and on exception clause approaches to nonmonotonic reasoning, which is not argument-based reasoning.

Although Brewka's PDL uses a priority treatment, which even allows defeasible priorities, the rules do not have weak antecedents, and can not express the difference between two defeasible rules, as was shown by Prakken (1997). But Brewka's Prioritized Extended Logic approach is important and valuable for this study, since it regards a logical priority treatment. For this study, priority is relevant in reasoning, as Brewka claims in his research, but the system is also argument-based, allowing defeasible treatment.

4.5 Other Researches

Other relevant researches have been done. I shall not discuss them in this paper, although they are of considerable importance

5 Modelling legal reasoning with argumentation and purposes

This is basically an Argumentation System approach, which uses nonmonotonic logic and defeasible reasoning. But this system expects to treat principles and rules separately.

[5] In which a collision could be won.

The open texture antecedent has the derivation acceptance by justifying its contents. The conflict of rules is treated by avoiding invalid rules or conceptual interpretations. The collision of principles is treated by weighing the purpose aimed at by the principle itself. So, it is priority oriented too. This priority can change, if the purposes change, modifying the conclusion, without any inconsistency to the system.

The internal structure of an argument has been defined. It is used to treat exceptions on a defeasible based system, which includes also the social mutation of the semantic and pragmatic meaning of a legal rule.

6 Conclusion

In this paper, relevant aspects of Law Argumentation Theory were discussed. They could have consequences on AI&Law, upon its purpose to offer logical tools for arising rationality in Legal Argument and Legal Discourse. A Legal Reasoning is a complex set of human beings, but it is important to model it too. We can not forget that rational discourse demands justification.

Some brief conclusions are: the classification of an open texture concept could be obtained, by justifying the contents of the concept; principles and rules have different behaviours in legal domain, specifically in conflicting situations; conflict of rules imply the invalidity of a rule or a meaning; collision of principles are treated by weighing the purpose aimed at by the principle; if the goal to be achieved by the action changes, the priority can change, thus, modifying the conclusion, without inconsistency to the system; the priority and open texture treatment ought to be integrated with argument-based reasoning, on account of the defeasible aspect of legal argument; the argument level should treat the semantic mutation of the concept, bringing up a pragmatic level. Other conclusions will only be obtained after discussion and after other researches.

Acknowledgements

The author wishes to thank Dr. Donald Peterson, who kindly accepted this paper on AISB 2000. The author also wants to thank FDV - Faculdades Integradas de Vitória, for supporting this study.

References

Alexy, Robert. (1978). Theorie der Jusistischen Argumentation. Die Theorie des rationalen Diskurses als eine Theorie der juristischen Begründung. Frankfurt am Main: Suhrkamp Verlag (the references are to the reprint).

Bench-Capon, T.J.M., (1993). Neural Networks and Open Texture. ACM Press.

Bench-Capon, T.J.M., Visser, P.R.S. (1996). Deep models, ontologies and legal knowledge based systems. In Foundations of legal knowledge systems. JURIX'96. Tilburg University Press

Dworkin, Ronald (1978) Taking Rights Seriously, Harvard University Press, 2 nd edicion.

Habermas, Jürgen. (1973). Legitimationsprobleme im Spätkapitalismus. Frankfurt am Main.

Habermas, Jürgen. (1973a). Wahrheitstheorien. In *Wirklichkeit und Reflexion. Festschrift für W. Schulz.* Pfullingen: H. Fahrenbach.

Hage, Jaap. (1995). Teleological Reasoning in Reason-Based Logic. ACM Press.

Hubmann, Heinrich. Festschrift für Schnorr-von Carolsfeld. In *Larenz, Karl. Methodenlehre der Rechtswissenschaft.* 5ª. ed. Berlin. Heidelberg. Lisboa: 1983.

Larenz, Karl. (1983) Methodenlehre der Rechtswissenschaft. 5ª. ed. Berlin. Heidelberg: Springer-Verlag. Lisboa: Fundação Calouste Gulbenkian.

Perelman, Chaïm/Olbrerechts-Tyteca, L. La nouvelle rhétorique, Traité de l'argumentation. Paris. 2 ª. ed. Bruxelles, 1970.

Prakken, Henry (1997) Logical Tools for Modelling Legal Argument. A Study of Defeasible Reasoning in Law. Netherlands

Rawls, John. (1971). A Theory of Justice. Cambridge: Cambridge University Press (the references are to the reprint).

Toulmin, Stephen E. (1958). The uses of argument. Cambridge: Cambridge University Press (the references are to the reprint).

Verheij, Bart. (1996) An integrated view on rules and principles. In JURIX'96 Foundations of legal knowledge systems. Paper version: Tilburg University Press.

Distributed Belief Revision as Applied Within a Descriptive Model of Jury Deliberations

Aldo Franco Dragoni,* Paolo Giorgini,** Ephraim Nissan***

* Istituto di Informatica, University of Ancona, Via Brecce Bianche, 60131 Ancona, Italy. dragon@inform.unian.it
http://www.inform.unian.it/personale/~dragon/dragoni.html

** DISA, Univ. of Trento, Via Inama 5, Trento, Italy. pgiorgini@cs.unitn.it

*** School of Computing and Mathematical Sciences, University of Greenwich, Wellington Street, Woolwich, London SE18 6PF, U.K.
E.Nissan@greenwich.ac.uk http://www.gre.ac.uk/~E.Nissan
http://www.gre.ac.uk/~E.Nissan/CmsWeb/ai_and_law_group.html

Belief revision is a well-research topic within AI. We argue that the new model of distributed belief revision as discussed here is suitable for general modelling of judicial decision making, along with extant approach as known from jury research. The new approach to belief revision is of general interest, whenever attitudes to information are to be simulated within a multi-agent environment with agents holding local beliefs yet by interacting with, and influencing, other agents who are deliberating collectively. In the approach proposed, it's the entire group of agents, not an external supervisor, who integrate the different opinions. This is achieved through an election mecahnism. The principle of "priority to the incoming information" as known from AI models of belief revision are problematic, when applied to factfinding by a jury. The present approach incorporates a computable model for local belief revision, such that a principle of recoverability is adopted. By this principle, any previously held belief must belong to the current cognitive state if consistent with it. For the purposes of jury simulation such a model calls for refinement. Yet, we claim, it constitutes a valid basis for an open system where other AI functionalities (or outer stimuli) could attempt to handle other aspects of the deliberation which are more specific to legal narratives, to argumentation in court, and then to the debate among the jurors.

1 Jury Research

Reid Hastie's paper collection, *Inside the Juror* (1993), is now a classic of jury research, and the psychology of judicial decision making by lay factfinders: descriptive models of juror decision making. Already in 1983, Hastie, Penrod and Pennington had published *Inside the Jury.* This domain has eluded thus far the mainstream of AI & Law research.

It appears to be the case that the very first paper published in an AI forum in the domain was Gaines et al. (1996). It described a neural model simulating juror decision making according to one of the several approaches current in psychologists' formal modelling of juror decision making.

Disciplines contributing to the approaches presented in *Inside the Juror* include "social psychology, behavioral decision theory, cognitive psychology, and behavioral modeling" (from the blurb on the back cover), yet this list is not complete. For example, in Ch. 11, Ehud Kalai sketched a game-theoretic framework. The affiliation of the authors in that volume is with schools of Law, departments of Statistics, or Psychology, or Management, or social or political science, but none comes from computer science. The discipline will have to take notice. Ours is a step in that direction.

Hastie's long introduction to his volume is usefully detailed—an excellent overview which we summarise below by way of introduction to our own application of a new general model of distributed belief revision. "One development in traditional jurisprudential scholarship is a candidate for the role of a general theory of juror decision making; namely the utilitarian model of rational decision making that has been imported into jurisprudence from economics" (4). Optimal decision making has been modeled, in the literature, not just for the role of the juror, but for the judge, attorney, police, and perpetrators of criminal behavior as well. Optimality, or rationality, for decision making is too strong an assumption (5). The research in *Inside the Jury* "focuses on the manner in which jurors behave before they enter the social context of deliberation in criminal felony cases" (5), with "at least four competing approaches represented" among behavioral scientists' descriptive models of decision making (10), namely, such that are "based on probability theory, 'cognitive' algebra, stochastic processes, and information processing theory" (10–11). Bayes' theorem is involved, in the former, for descriptive purposes in *Inside the Juror*—being applied to the psychological processes in which a juror is engaged—rather than in prescribing how to evaluate evidence to reach a verdict, "or to evaluate and improve jurors' performance" (12).

Note that Bayesianism in legal evidence research is a controversial, hotly debated topic: Allen and Redmayne (1997) is a journal special issue with contributions from both camps, namely the Bayesio-skeptics and the so-called Bayesian enthusiasts. Yet, when it comes to descriptive models of how jurors shape their opinions, it's not obvious *prima facie* that the controversy extends into jury research. Jurors do *not* reason about the evidence according to the Bayes theorem, it may be argued, and even if they tried to apply probability explicitly, they would lack the formal skills to do so. This is beyond the point. Rather, among the descriptive models of juror decision making there are *also* probabilistic or stochastic models, to describe a process in general terms—not for a specific case at hand. This point is essential for making sense of our contribution in this paper.

The second class of approaches to juror's decision making, as enumerated in the introduction to Hastie's volume, fits among such psychological theories of mental processes that are couched in the form of algebraic equations (17), with evidence being combined according to a weighted average equation. "As in the Bayesian model, we are dealing with a single meter in which the results of all the subprocesses are summarized in a current belief and in which the ultimate 'categorical' verdict decision is based on the comparison of the final belief meter reading to a threshold to convict" (19), but belief updating in the algebraic approach is additive instead of multiplicative as in Bayesian models, and moreover extreme judgments are adjustable instead of final.

Stochastic process models are the third family; they differ from the previous two in that the larger process is assumed to behave in a random fashion, and what is probabilistic is state transitions over time. The fourth family adopts the information processing paradigm from cognitive psychology; they are typified by the room they make for mental representations, memory activation, elementary information processes, an executive monitor, and a specific cognitive architecture.

For example, Reid Hastie's Ch. 4 in his volume is devoted to algebraic models: the basic averaging and the sequential averaging models. "To date, the most visible accomplishments have been byproducts of the algebraic application; e.g., useful individual-level numerical indices of the importance of evidence, presumptions of innocence, and standards of proof" (110). Norbert Kerr (Ch. 5) is concerned with stochastic models; David A. Schum and Anne W. Martin (Ch. 6), with probabilistic evidence probativity assessment. Schum, by background a psychologist who has also researched at the meet of computing and operations research as well as law, is indeed one of the most visible representatives of the Bayesian camp within legal evidence research, yet he has also made other important contributions—especially his adaptation of Wigmorean analysis, as well as his repertoire of basic formal op[erations for marshalling the evidence—whose value is fairly acceptable also for Bayesio-skeptics and can arguably be embraced by AI & Law research with little risk of antagonising those sceptical about Bayesianism's value for the analysis of the evidence in a given legal case.

Schum's and Martin's chapter in Hasdtie's volume has a major focus "upon inductive inference tasks, which Wigmore [1937] termed 'catenated'; the modern terms for these tasks are 'cascaded' or 'hierarchical' " (136). In contrast Ch. 7, by Schum, which eventually also applies Bayesian likelihood-ratio formulations for weighing evidence, places more emphasis on argument structuring. Arguably, this could be an entry point into the domain for such AI & Law researchers whose interests are in argumentation models. In Ch. 8, Nancy Pennington and Reid Hastie present a cognitive theory of story construction on the part of the juror, and indeed we propose (see Nissan's paper on the JAMA model in this forum) that models of narrative understanding from natural-language processing are all-important, if one is to apply, next, AI & Law to narratives of a case at hand or, perhaps preferably, to narrative patterns for siituational classification purposes within problem-solving tasks.

This is the backdrop from jury research, for our proposed application of the new model of distributed belief revision based on the principle of recoverability as explained at the start.

2 A Novel General Approach

Jurors' opinions and beliefs are destined to evolve as the trial goes on. New information and evidence integrate and corroborate the cognizance of the Court, but other testimonies might cause conflicts. In this case, it seems natural that the acquisition of the new evidence s hould be accompanied by a reduction of the credibility of the conflicting pieces of knowledge. If the juror's corpus of evidence is not a flat set of facts but contains rules, finding such conflicts and determining all the sentences involved in the contradictions can be hard. In dealing with these "changes of mind" we heavily relies on symbolic logic, since as much as it contributed to the history of "thinking", logic could as well solve the problem of "thinking over". AI reserchers call this cognitive process "belief revision".

Since the seminal, philosophical and influential works of Alchourrón, Gärdenfors and Makinson (1985) ideas on "belief revision" have been progressively refined (Gärdenfors 1988) toward normative, effective and computable paradigms (Benferhat et al. 1993; Nebel 1994). They introduced three rational principles to whom belief revision should obey:

AGM1 *Consistency*: revision must yield a consistent knowledge space.

AGM2 *Minimal Change*: revision should alter as little as possible the knowledge space.

AGM3 *Priority to the Incoming Information*: incoming information always belongs to the revised knowledge space.

They conceived a cognitive state K as a deductively closed set of sentences of a formal language L. From $AGM1 \div AGM3$ they drew up eight postulates for belief revision. Here K^*p denotes the cognitive state K revised in the light of the incoming information p, while K^+p denotes the deductive closure of $K \cup \{p\}$.

K^*1. For each p and K, K^*p is still a cognitive state
K^*2. $p \in K^*p$
K^*3. $K^*p \subseteq K^+p$
K^*4. If $\neg p \notin K$ then $K^+p \subseteq K^*p$
K^*5. K^*p is inconsistent iff p is inconsistent
K^*6. If p and q are logically equivalent then $K^*p = K^*q$
K^*7. $K^*(p \wedge q) \subseteq (K^*p)^+q$
K^*8. If $\neg q \notin K^*p$ then $(K^*p)^+q \subseteq K^*(p \wedge q)$

These axioms describe the rational properties to which revision should obey, but they do not suggest how to perform it. An obvious way is that of deleting $\neg p$ from K (reducing in some way K at a point that $\neg p$ is no longer derivable), adding p and making the deductive closure The deletion of $\neg p$ from K, $K^-\neg p$, is called contraction, and can be defined in terms of "Epistemic Entrenchment" (Gärdenfors 1988), which is an ordering $\leq$, that envisages the logical dependencies of the formulae in K; it depends on K but it applies to all the formulae of L. $p \leq q$ means that p is less entrenched (i.e., more exposed to eventual changes) than q. $\leq$ satisfies the following postulates:

EE1. $\leq$ is transitive
EE2. For all p, $q \in L$, if $p \vdash q$ then $p \leq q$
EE3. For all p, $q \in L$, either $p \leq p \wedge q$ or $q \leq p \wedge q$
EE4. If K is consistent, then $p \notin K$ iff for all $q \in L$, $p \leq q$
EE5. If for all q of L, it holds $q \leq p$, then p is a tautology

Contraction could be defined from the Epistemic Entrenchment as follows: $q \in K^-p$ iff $q \in K$ and, either $p < q \vee p$, or p is a tautology. K^-p contains only the formulae of K that have a greater degree of epistemic entrenchment than p. There are three problems with such a kind of revision:

1. it deals with infinite sets of sentences
2. $\leq$ depends on K, so it is difficult to iterate the revision because the ordering defined on K^*p could be different from the one defined on K
3. the choice of a particular ordering $\leq$ satisfying the postulates EE1$\div$ EE5 is arbitrary; as Gärdenfors (1988) wrote: "[the postulates] leave the main problem unsolved: what is a reasonable metric for comparing different epistemic states?".

Indeed, regarding the latter problem, one of the claim of this paper is that, *such computable and reasonable metric can be provided only by numerical approaches*. The AGM approach to belief revision do respect Dalal's (1988) "principle of irrelevance of the syntax" by which, syntactically different but logically equivalent formulae represent the same knowledge space. The partisans of *syntax-dependent* belief revision consider knowledge spaces made up of a limited number of sentences. They claim that asserting facts is more important than deriving others from them. Nebel's (1994) *epistemic relevance ordering* stratifies a base B into n priority classes $B_1, ..., B_n$. Epistemic relevance does not respect the logical contents of the sentences as epistemic and partial entrenchment do. A justification seems to rely on the logical paradoxes of the material implication: a rule $q \rightarrow p$ should not

necessarily be considered more important than p just because $p \vdash q \rightarrow p$. Let $B{\downarrow}p$ denote the set of the subsets of B that fail to imply p. Nebel defines $B{\Downarrow}p$ as the subset of $B{\downarrow}p$ made of the elements that *contain as many sentences of the highest priority as possible.*

The corresponding revision is defined as:

$$B \oplus p = Th\left(\left(\bigcap_{B' \in (B \Downarrow \neg p)} Th(B')\right) \cup \{p\}\right)$$

where $Th(B')$ denotes the deductive closure of B'. There are two problems with this revision: =

- it does not satisfy all the AGM postulates
- it is still computationally hard.

We could adopt various criteria to sort and select the elements of $B{\downarrow}p$. Let $B' = B'_1 \cup ... \cup B'_n$ and $B'' = B''_1 \cup ... \cup B''_n$ two consistent subsets of B where $B'_i = B' \cap B_i$ and $B''_i = B'' \cap B_i$. Benferhat et al. (1993) [cf. Dubois & Prade (1992)] suggest (implicitly) three ways

to translate the epistemic relevance into a preference relation $\leq$ on $B \downarrow p$.

- *best-out ordering.* $B'' \leq B'$ iff the most credible of the sentences in $B \setminus B''$ is less credible than the most credible of the sentences in $B \setminus B'$.
- *inclusion-based ordering.* $B'' \leq B'$ iff there exists a stratum i such that $B'_i \supset B''_i$ and for any $j < i$, $B'_j = B''_j$. This preordering is strict but partial; its maximal consistent elements are also maximal for the best-out ordering.
- *lexicographic ordering.* $B'' \leq B'$ iff there exists a stratum i such that $|B'_i| > |B''_i|$ and for any $j < i$ $|B'_j| = |B''_j|$, and $B'' = B'$ iff for any j, $|B'_j| = |B''_j|$.

$B{\Downarrow}p$ contains the elements of $B{\downarrow}p$ maximal w.r.t. inclusion-based ordering.

A juror's cognitive state does not suffer only from inconsistency; it can also be affected by uncertainty. Numerical distributions of credibility over the sentences of L or over the set Ω of the models of L, play the same role that "epistemic entrenchment", p.e.r. and epistemic relevance play in the symbolic frameworks. Generally, numerical approaches do not respect logical dependencies among the sentences. Logics of uncertainty often represent a cognitive state K and the incoming information p in terms of their sets of models (also said "possible worlds"), respectively, $[K]$ and $[p]$. A cognitive state is

represented not simply by $[K]$, but by an assignment function $d(\omega) : \Omega \rightarrow [0, 1]$ such that $d(\omega') = 0$ for each $\omega' \notin [K]$. The arrival of p generally means that the real world belongs to $[p]$. This event changes d into a new assignment (new prioritization) d'. Imposing the priority to the incoming information (AGM3) means assigning $d'(\omega) = 0$ to each $\omega \notin [p]$. Minimizing this change (AGM2) means minimizing some kind of distance between d and d'.

In the probabilistic approach (Pearl 1988) a cognitive state is characterized by a *probability measure* P on 2^Ω, whose fundamental property is *additivity*: $\forall A, B \subseteq \Omega, A \cap B \neq \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$. $P(\Omega) = 1$, so if $\bar{A} = \Omega - A$ then $P(A) + P(\bar{A}) = 1$. We might also consider the *probability distribution* $pr(\omega)$ that assignes a probability degree to each world in Ω. where

$$P(A) = \sum_{\omega \in A} pr(\omega) \ .$$

$pr(\omega) = 0$ means that ω is not a possible world. $pr(\omega) = 1$ means that ω is surely the real world. An incoming information p changes the probability measure of any sentence q of L through the very famous Bayes' Conditioning Rule:

$$P(q|p) = \frac{P([q] \cap [p])}{P([p])} = \frac{P([p]|[q]) \cdot P([q])}{P([p])}$$

which can also be expressed in terms of probability distribution:

$$pr(\omega|p) = \begin{cases} \dfrac{pr(\omega)}{P([p])} & \text{if } \omega \in [p] \\ 0 & \text{otherwise} \end{cases}$$

This modification is defined only for $P([p]) > 0$, hence it is not applicable when p is judged impossible by the previously determined probability measure P. =

Bayesian conditioning obeys the principle of priority to incoming information (AGM3); it increases the probability of the not-impossible worlds belonging to $[p]$ to the prejudice of those external to $[p]$ which become all impossible. =

In the probabilistic framework the probability of a sentence p is simply the probability measure $P([p])$. Thus, probability measures order the sentences of L, but, unfortunately, they do not generate epistemic entrenchments. In effect, probability measures satisfy EE1 since they are, obviously, transitive (if $P([p]) \leq P([q])$ and $P([q]) \leq P([r])$ then $P([p]) \leq P([r])$. EE2 too is verified since $p \vdash q$ means $[p] \subseteq [q]$ hence $P([p]) \leq P([q])$ (it is always easier to retract p than q). Even EE4 is verified; in fact, $p \notin K$ means $P([p]) = 0$, and if K is consistent then there are sentences q such that $P([q]) = 0$, hence $P([p]) = 0$ iff $\forall q \in L$ $(P([p]) \leq P([q]))$. Finally, EE5 is verified since if $\forall q \in L$ $P([q]) \leq P([p])$ then $[p] = \Omega$ which means that p is a tautology. Unfortunately, EE3 is generally unsatisfied since $[p \wedge q] \subseteq [p]$ and $[p \wedge q] \subseteq [q]$ so that $P([p \wedge q]) \leq P([p])$ and $P([p \wedge q]) \leq P([q])$; normally it is easier to retract a conjunction than any of its conjuncts.

Also the belief function framework (Shafer 1990; Shafer & Srivastava 1990) assigns a probability P to the subsets of $= \Omega$, with the constraints $P(\emptyset) = 0$ and $\sum_{A \subseteq \Omega} P(A) = 1$. If $P(A) > 0$ then A is said to be a *focal element*. The *belief function* on the subsets of Ω is defined as

$$Bel(A) = \sum_{X \subseteq A} P(X)$$

$Bel(A)$ measures the persuasion that the real world is inside A; maybe that there is no evidence that directly support A but it cannot be excluded because there is evidence that supports some of its subsets. This function is not additive: $Bel(A) + Bel(\bar{A}) \leq 1$. The knowledge is:

- *certain* and *precise* if there exists a $\omega \in \Omega$ such that $P(\{\omega\}) = 1$
- *certain* and *imprecise* as if there exists an $A \subset \Omega$ such that $P(A) = 1$ but A is not singleton
- *consistent* if all the focal elements are nested
- *inconsistent* if all the focal elements are disjoint
- *void* if $P(\Omega) = 1$ and for all $A \subseteq \Omega$, $P(A) = 0$.

This framework deals also with uncertain inputs. They are treated as new probability assignments on 2^{Ω}. The change consists of merging the two evidences (the prior P_1 and the new P_2) through the Dempster's Rule of Combination:

$$P(A) = \frac{\sum_{X_1 \cap X_2 = A} P_1(X_1) \cdot P_2(X_2)}{\sum_{X_1 \cap X_2 = \emptyset} P_1(X_1) \cdot P_2(X_2)}$$

for all $A \subset \Omega$. This rule, easily extensible to combine n probability assignments, reinforces concordant evidence and weakens conflicting ones. It can be applied only if evidences are independent and referred to the same Ω. Because of the commutativity of the product, the rule is independent from the sequence $P_1 ... P_n$ so *it violates the principle of priority to the incoming information*! From a knowledge engineering point of view, the worst problem with the Dempster's Rule of Combination is its computational complexity. One should generate a frame of $2^{|\Omega|}$ elements to calculate it! However, much work has been spent in reducing the complexity of that rule. Such methods range from "efficient implementations" (Kennes 1992) to "qualitative approaches" (Parson 1994) through "approximate techniques" with statistical methods as the Montecarlo sampling algorithm (Wilson 1991; Moral & Wilson 1996).

3 Requirements for a Belief Revision Framework in a Multi Source Environment

We think that to revise beliefs in a Multi-Agent scenario, where many sources give information about a same static situation, the framework should satisfy some requisites.

- Ability to reject incoming information

Jurors should not obey the principle of "priority to the incoming information" which is not acceptable since there is no strict correlation between the chronology of the informative acts and the credibility of their contents (Dragoni, Mascaretti & Puliti 1995); it seems more reasonable to treat all the available pieces of information as they had been collected at the same time.

- Ability to recover previously discarded beliefs

Jurors should be able to recover previously discarded pieces of knowledge after that new evidence redeems them. The point is that this should be done not only when the new information directly "supports" a previously rejected belief, but also when the incoming information indirectly supports it, by disclaiming the beliefs that contradicted it, causing its ostracism. More formally, for each cognitive state K, and sentences p and q such that $K \vdash p$ and $K^*q \nvdash p$, there can always be another piece of information r such that $(K^*q)^*r \vdash p$, even if $r \nvdash p$. An obvious case should be $r = \neg q$. We elsewhere called this rule *principle of recoverability*: "any previously held piece of knowledge must belong to the current knowledge space if consistent with it" (Dragoni, Mascaretti & Puliti 1995; Dragoni 1997; Dragoni & Giorgini 1997a).

The rationale for this principle is that, if someone gave us a piece of information (sometime in the past) and currently there is no reason to reject it, then we should accept it! This is stronger than the traditional "coherence" spirit of belief revision, since the piece of knowledge to accept is not a *generic* sentence of the language but a *generated* piece of information; somewhere there is an utilitarian intelligent information source that guarantees for it. Of course, this principle does not hold for updating, where changes may be irrevocable. This feature could also be subtitled: "revocable treatment of consistency" . We remember of Minsky's lection: "I do not believe that consistency is necessary or even desirable in a developing intelligent system ... What is important is how one handles paradoxes or conflicts ... *Enforcing consistency produces limitations*. As we will see in a moment, we overcome this problem by defining a single global,

never forgetting, eventually inconsistent *Knowledge Background*, upon which act multiple specific, competitive, ever changing, *consistent* cognitive states.

- Ability to *combine* contradictory and concomitant evidences

The notion of *beliefs integration* should blend that of revision (Dragoni & Giorgini 1997b). Every incoming information changes the cogniti= ve state. Rejecting the incoming information does not mean leaving beliefs unchanged since, in general, incoming information alters the distribution of the weights. Surely the last incoming information decreased the credibility of the beliefs with whom it got in contradiction, even in the case that it has been rejected. The same when receiving a piece of information which we were already aware of; it is not the case that nothing happened (as AGM K^*4 states) since we are now, in general, more sure about that belief. More generally, there is no reason to limit the changes introduced by the new information to an insertion into a pre-established relative order with consequent rearrangement of the ranking to accomplish the logical relations between beliefs (as Williams' transmutation does). If it is true that new incoming information affects the old one, it is likewise true that the latter affects the former. In fact, an autonomous agent (where "autonomous" means that his cognitive state is not *determined* by other agents) judges the credibility of new information on the basis of its previous cognitive state. "Revising beliefs" should simple mean "dealing with a new broader set of pieces of information".

- Ability to deal with couples <source,information> rather than with information alone

The way the credibility ordering is generated and revised must reflect the fact that beliefs come from different sources of information, since the reliability and the number of independent informants affect the credibility of the information and vice versa (Dragoni 1992).

- Ability to maintain and compare multiple candidate *cognitive states*

This ability is part of humans intelligence which does not limit its action to comparing single pieces of information but goes on trying to reconstruct alternative cognitive scenarios as far as it is possible.

- Sensibility of the syntax

Despite Dalal's (1988) aforementioned principle, syntax plays an important role in everyday life. The way we pack (and unpack) pieces of information reflects the way we organize thinking and judge credibility, importance, relevance and even truthfulness. A testimony of the form $\alpha \wedge \delta \wedge ... \wedge \zeta \wedge \neg\alpha$ from a defendant A in a trial has the same semantic truth value than the testimony $\beta \wedge \neg\beta$ from defendant B, but we remember many cases in which B has been condemned while A has been absolved, being regarded his/her testimony "partially true", contrasting with the B's one regarded as "absolutely contradictory". A set of sentences seems not to be logically equivalent to their conjunction and we could change a cognitive state by simply clustering the same beliefs in a different way.

4 A Computable Model for Belief Revision

Our sentence-based approach for belief revision (Dragoni 1997) envisages two knowledge repositories:

1. the *knowledge background* KB, which is the set of all the propositional sentences available to the reasoning agent (as assumptions); it can be inconsistent

2. the *knowledge base* $B \subseteq KB$, which is the maximally consistent, currently preferred piece of knowledge that should be used for reasoning and decision supporting

Computationally, our way to belief revision consists of five steps (Dragoni & Giorgini 1997a,b):

S1. detection of the minimally inconsistent subsets of $KB \cup \{p\}$ (*nogoods*)
S2. generation of the maximally consistent subsets of $KB \cup \{p\}$ (*goods*)
S3. revision of the credibility weights of the sentences in $KB \cup \{p\}$
S4. choice of a preferred *good* as the new revised base B'

S5. selection of the derived sentences which are derivable from B'

The incoming information p, with its weight of evidence, is confronted not just within the current base B, but within the overall knowledge background KB. Doing so, the degrees of credibility of the sentences in $KB \cup \{p\}$ are reviewed on a broader and less prejudicial basis (S3). As already explained, the main advantage is that we can rescue sentences from KB by virtue of the maximal consistency of B'. If we'd revise only B by p, we could not recover information from KB. For instance, Nebel's revision would select

some $B' \in B\Downarrow\neg p$, but it will be always possible to find out some $B'' \in KB\Downarrow\neg p$ such that $B' \subseteq B''$.

S4 might choose a new base B' syntactically equal to the previous B (meaning that p has been rejected) but, in general, B' will have a different credibility distribution than B. p might be rejected even if S4 chooses a base B' different from B, but that still containing sentences incompatible with p.

When p is consistent with B, not necessarily $B'=B\cup\{p\}$, since S3 may yield a totally different choice at S4. Previously rejected pieces of knowledge $R \subset KB$ can be rescued simply by determining some upsetting between the credibility of a set $S \subset B$ and the credibility of R, this may happen if p *supports* R against S. The rejection of the priority to the incoming information principle implies that K^*4 and and K^*5 hold no longer (if p is inconsistent it will be part of none of the goods produced at S2, so it will never be part of a base).

S1, S2 and S5 deal with consistency and derivation, and act on the symbolic part of the information. Operations are in ATMS style; to find out nogoods and goods, we adopt (and adapt) the most efficient set-covering algorithm that we are aware of Reiter (1987). Notwithstanding this, even in the propositional case, determining all the minimal inconsistencies can be very hard. However, such condition can be relaxed (the consequence is that some of the goods are not really consistent) and in practical applications dealing with commonsense knowledge (see e.g. Dragoni & Di Manzo 1995), such minimal inconsistencies could be provided interactively from the outside by the user.

S3 and S4 deal with uncertainty and work with the numerical weight of the information. Both contribute to the choice of the revised knowledge space so their reasonableness should be evaluated as a couple. Numerical formalisms are able to perform both of them since the credibility of a single sentence p is determined in the same way as the credibility of a set of sentences B by the weights attached to $[p]$ and $[B]$, respectively. Flexibility is an advantage in separating the two steps; for instance, depending on the characteristics of the knowledge domain under consideration and the kind of task and/or decision that should be taken on the basis of the revision outcome, the selection function could consider also one (or a combination) of the methods described in Benferhat et al. (1993).

Probabilistic methods with uncertain inputs seem inadequate for the strong dependence that they impose on the credibility of a sentence and that of its negation. We see that the belief-function formalism, in the special guise in which Shafer and Srivastava (1990) apply it to auditing, could work well because it treats all the pieces of information as they had been provided at the same time.

The method has the following I/O (see Dragoni & Giorgini 1997a):

Input:

list of pairs <source,piece of information>
list of pairs <source,reliability>

Output:

list of pairs <piece of information,credibility>
list of pairs <source,reliability>

Let S=$\{s_1, ..., s_n\}$ be the set of the sources, and let kb_i be the subset of KB received from s_i. Each source s_i is associated with a *reliability* $R(s_i)$, that is regarded as the *probability* that the source is faithful. The main idea with this multi-source version of the belief function framework is that a reliable source cannot give false information, while an unreliable source can give correct information; the hypothesis that s_i is reliable is compatible only with the models of kb_i, while the hypothesis that s_i is unreliable is compatible with the overall Ω. Each source s_i is an evidence for KB and generates the following *bpa* $m_i(\cdot)$ on 2^Ω:

$$m_i(X) = \begin{cases} R(s_i) & \text{if } X = [kb_i] \\ 1 - R(s_i) & \text{if } X = \Omega \\ 0 & \text{otherwise} \end{cases}$$

All these *bpas* will be then combined through the Dempster Rule of Combination. From the combined *bpa* $m(\cdot)$, the credibility of a sentence p of L is given, as usual, by:

$$Bel(p) = \sum_{X \subseteq [p]} m(X)$$

From this mechanism we obtained an easy way to calculate the new reliabilities of the sources. Let Φ be an element of 2^S. If the sources are independent, the reliability of Φ is

$$R(\Phi) = \prod_{S \in \Phi} R(s) \cdot \prod_{S \in \Phi} (1 - R(s))$$

It holds that

$$\sum_{\Phi \in 2^S} R(\Phi) = 1$$

It may be that some source fall in contradiction, so that some elements of 2^S are impossible. The remaining elements are subjected to Bayesian conditioning so that their reliabilities sum up again to 1. The revised reliability $R^*(s)$ of a source s is the sum of the new reliabilities of the surviving elements of 2^S that contain s. If a source has been involved in

some contradictions, then $R^*(s) \leq R(s)$, otherwise $R^*(s) = R(s)$.

S4 translates such ordering on the *sentences* in $KB \cup \{p\}$ into an ordering on the *goods* of $KB \cup \{p\}$. The best classified good is selected as the preferred revised knowledge base. If the ordering on $KB \cup \{p\}$ is not strict, then there can be multiple preferred goods. In this case we could take their intersection as revised knowledge base (Benferhat et al. 1993). Yet, the intersection is not maximally consistent and this means that all the conflicting pieces of knowledge with the same credibility will be rejected.

Another question is: S4 should consider only the *qualitative* ordering of the sentences in $KB \cup \{p\}$ (relative classification without the numerical weights) or could it take advantage of the *explicit* ordering (numerical weights). The first approach seems closer to the human cognitive behavior (which normally refrains from numerical calculus). The second one seems more informative (it takes into account not only relative positions but also the gaps between the items). In our model we do not use the "best-out" ordering for its "drowning effect" (Benferhat et al. 1993). The lexicographic one could be justified in some particular application domains (e.g. diagnosys). The inclusion-based method seems the most reasonable since it eliminates always the least credible one among conflicting pieces of knowledge.

As an example of a numerical way to perform S4, ordering the goods according to their average credibility seems reasonable and easy to calculate. With this method the preferred good may not contain the most credible sentence.

In the belief function framework, a "good" g is an element of Ω, precisely the one in which all the sentences in g are considered "true" and all the sentences out of g are considered "false" . This implies that the belief-function formalism is able to attach directly a degree of credibility to g, bypassing S4 in our framework. Unfortunately, when a good contains only part of the information supplied by a source, the belief-function formalism puts at zero its degree of credibility. This is unreasonable and, unluckily, the event is all but infrequent, so that often the credibility of all the goods is null.

A final step in our revision mechanism is the selection of the derived sentences which are still derivable from B` since the assumptions on which they rely are all contained in $B$`. Theoretically, it simply consists in applying classical entailment on the preferred good to deduce plausible conclusion from it. We adopted an ATMS and we stored each sentence derived by the Theorem Prover with an *origin set* (Martins & Shapiro 1988), i.e., a set of basic assumptions which are all *necessary* to derive it. Practically, this step consists in selecting from the derived sentences, all those whose origin set is subset of the preferred good. We could relax the definition of origin set to that of a set of basic assumptions used to derive the sentence. This is easier to compute and does not have harmful consequences; the worst it can happen is that, being this relaxed origin set a superset of the real one, it is not certain that it will be a subset of the preferred good as the real one is, and so some derived logical consequences of the preferred good may be not recognized (at first).

Besides recoverability, this computational model for belief revision overcomes various limitations of other classic approaches, in particular:

- the revision can be iterated
- inconsistent incoming information does not yield inconsistent revised knowledge spaces
- the numerical revision is performed on a broader base (the overall KB)
- the revision is more flexible;
- the complete numerical ordering renders the revision as least drastic as possible
- the splitting between the symbolic treatment of the inconsistencies and the numerical revision of the credibility weights, provides a clear understanding of what is going on and lucid explanations for the choices.

Dragoni and Giorgini are currently applying this conception of belief revision in a distributed monitoring system (Dragoni & Giorgini 1998) and in the police inquiry domain (Dragoni, Ceresi & Pasquali 1996).

Within jury research, such a model of deliberative negotiation on opinion are not to be adopted "as is", as the model is likely to require fine-tuning to the specifics of trial contexts, let alone taking account of the exclusionary rules of evidence as reflected in the judge's instructions to the jury. Yet, arguably we have here an important approach that could eventually stand at least on a par with the approaches (especially the probabilistic or stochastic ones) represented in Hastie's volume.

References

C.E. Alchourrón, P. Gärdenfors, and D. Makinson (1985) "On the logic of theory change: Partial meet contraction and revision functions". *The Journal of Symbolic Logic*, 50: pp. 510–530.

R. Allen and M. Redmayne, eds. (1997) *Bayesianism and Juridical Proof.* Special issue of *The International Journal of Evidence and Proof*, 1. London: Blackstone. [Vol. 1 includes 4 regular, numbered issues and one thematic issue, unnumbered.]

S. Benferhat, C. Cayrol, D. Dubois D., J. Lang, and H. Prade (1993) "Inconsistency management and prioritized syntax-based entailment". *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, pp. 640–645.

M. Dalal (1988) "Investigations into a theory of knowledge base revision". *Proceedings of the 7th National Conference on Artificial Intelligence*, pp. 475–479.

A.F. Dragoni (1992) "A model for belief revision in a multi-agent environment". In: Werner E. and Demazeau Y., eds., *Decentralized AI. 3.* Amsterdam: North Holland / Elsevier Science.

A.F. Dragoni (1997) "Belief revision: from theory to practice". *The Knowledge Engineering Review*, 12(2).

A.F. Dragoni, C. Ceresi, and V. Pasquali (1996) "A system to support complex detective inquiries". *Proceedings of the Fifth Iberoamerican Conference on Computer Science and Law*, La Habana.

A.F. Dragoni and P. Giorgini (1997a) "Belief revision through the belief function formalism in a multi-agent environment". In: Wooldridge M., Jennings N.R., and Muller J., eds., *Intelligent Agents, III* (LNCS, vol. 1193.) Heidelberg: Springer-Verlag.

A.F. Dragoni and P. Giorgini (1997b) "Distributed knowledge revision-integration". *Proceedings of the Sixth ACM International Conference on Information Technlogy and Management* . New York: ACM Press.

A.F. Dragoni and P. Giorgini (1998) "Sensor data validation for nuclear power plants through bayesian conditioning and dempster's rule of combination". *Computers and Artificial Intelligence*, 17(2/3): pp. 151–168.

A.F. Dragoni and M. Di Manzo (1995) "Supporting complex inquiries". *International Journal of Intelligent Systems*, 10: pp. 959–986.

A.F. Dragoni, F. Mascaretti, and P. Puliti (1995) "A generalized approach to consistency-based belief revision". In: M. Gori and G. Soda, eds., *Proc. of the Conference of the Italian Association for Artificial Intelligence* (LNAI, vol. 992.) Heidelberg: Springer Verlag.

D. Dubois and H. Prade (1992) "Belief change and possibility theory". In: P. Gärdenfors, ed., *Belief Revision.* Cambridge University Press.

D.M. Gaines (1994) "Juror Simulation". BSc Project Report (Proj. No. CS-DCB-9320), Computer Science Dept., Worcester Polytechnic Institute.

D.M. Gaines, D.C. Brown and J.K. Doyle (1996) "A Computer Simulation Model of Juror Decision Making". *Expert Systems With Applications* 11(1): pp. 13–28.

P. Gärdenfors (1988) *Knowledge in Flux: Modeling the Dynamics of Epistemic States.* Cambridge, MA: MIT Press.

R. Hastie, ed. (1993) *Inside the Juror: The Psychology of Juror Decision Making* (Cambridge Series on Judgment and Decision Making.) Cambridge, U.K.: Cambridge University Press, 1993 (hc), 1994 (pb).

R. Kennes (1992) "Computational aspects of the Möbius transform of a graph". *IEEE Transactions in Systems, Man and Cybernetics*, 22: pp. 201–223.

J. De Kleer (1986) "An assumption based truth maintenance system". *Artificial Intelligence*, 28: pp. 127–162.

J.P. Martins and S.C. Shapiro (1988) "A model for belief revision". *Artificial Intelligence*, 35: pp. 25–97.

S. Moral and N. Wilson (1996) "Importance sampling monte-carlo algorithms for calculation of Dempster-Shafer belief". *Proceedings of IPMU'96,* Granada, Spain.

B. Nebel (1994) "Base revision operations and schemes: Semantics, representation, and complexity". In: A.G. Cohn, ed., *Proceedings of the 11th European Conference on Artificial Intelligence.* John Wiley & Sons.

S. Parson (1994) "Some qualitative approaches to applying the Dempster-Shafer theory". *Information and Decision Technologies*, 19: pp. 321–337.

J. Pearl (1988) *Probabilistic Reasoning for Intelligent Systems.* San Mateo, CA: Morgan Kaufmann.

R. Reiter (1987) "A theory of diagnosis from first principles". *Artificial Intelligence*, 53.

G. Shafer (1990) "Belief functions". In: G. Shafer and J. Pearl, eds., *Readings in Uncertain Reasoning.* San Mateo, CA: Morgan Kaufmann.

G. Shafer and R. Srivastava (1990) "The Bayesian and belief-function formalisms: a general perspective for auditing". In: G. Shafer and J. Pearl, eds., *Readings in Uncertain Reasoning.* San Mateo, CA: Morgan Kaufmann.

N. Wilson (1991) "A Monte-Carlo algorithm for Dempster-Shafer belief". In: P. Smets, B.D. D'Ambrosio and P. Bonissone, eds., *Proceedings of the Seventh Conference*, pp. 414–417.