Criminal Sentencing and Intelligent Decision Support

URI J. SCHILD
Department of Mathematics and Computer Science, Bar-Ilan University, Ramat-Gan 52900, Israel
E-mail: schild@bimacs.cs.biu.ac.il

Abstract. The construction of computer systems for sentencing decision support has lately been of increasing interest. Different approaches and models exist for such systems. Of the available technologies we advocate improved access to precise statistical data, reject a rule-based expert system approach, and argue that a case-based advisory system approach could be the most acceptable. Two prototype systems we have under development and testing are described.

> "There are two kinds of people in prison: those who should never have been sent there, and those who should never be let out."

> (George Bernard Shaw)

1. Introduction

1.1. THE PROBLEM

The problem described here deals with the question how computer systems may assist judges in the process of passing sentence in criminal cases. As should seem obvious the sentencing process calls for human discretion: Apart from mandatory sentences for specific offences, judges are usually faced with having to choose one out of many acceptable sentences. This process of selection and decision malting is a most distinctive human task, and attempts at its computer simulation would form one of the major challenges for Artificial Intelligence.

The present state-of-art in AI does not enable such a simulation. As a matter of fact, most people – computer scientists, legal experts and laymen – do not want to consider the development of computer systems that independently would carry out the tasks of a judge. They believe that as a matter of *principle* this activity should be an exclusively human one (see (Gardner 1987, 79–83) for a fuller discussion of this issue, and other problems relating to misuse of computer programs in law).

The goal should rather be to build a decision *support* system for sentencing. Such a system would supply a judge with relevant information and advice. The operation of the system could be based on statistical or criminological models, and

it could also involve an expert knowledge-base. Information would be retrieved by the system through an intelligent reasoning process. But the final decision would be made by the judge himself.[1]

The construction of such a system is also a major challenge. It is our goal to discuss what kind of support a computer system can give to a judge with respect to the sentencing process. We shall consider what other researchers have done on this subject, and also describe our own work. We shall make a distinction between non-intelligent computer systems, which are essentially statistical or criminological in nature, and expert systems, i.e., systems which exhibit intelligence of their own.

The concept of an expert system does not have an exact definition anymore (if it ever had one). We shall call the kind of system that passes sentence on its own, an expert system, while the kind of system we are interested in could be called an intelligent advisory system.

The discussion of our problem is also relevant to similar problems in other areas. Determination of amounts of child support and alimony, or calculation of reparations for physical injuries are problems where human discretion is applied to fixing a numerical value. There may also be similar problems outside the legal domain.

1.2. PRACTICAL APPLICATION

The topic of legal computer systems for sentencing is of interest both from a theoretical jurisprudential point of view, and a criminological one. It is of course of relevance for computer scientists working in the area of AI and Law, and it can also have a major practical importance, as we shall now explain.

The main reason for developing and applying sentencing advisory systems relates to the quality of the sentencing process as viewed by the public at large and by the judiciary itself. Consider the following story by John Dewey as quoted in (Gardner 1987, 83):

> There is the old story of the layman who was appointed to a position in India where he would have to pass in his official capacity on various matters in controversy between the natives. Upon consulting a legal friend, he was told to use his common-sense and announce his decisions firmly; in the majority of cases his natural decision as to what was fair and reasonable would suffice. But his friend added: Never try to give reasons, for they will usually be wrong. (Dewey 1924, 17)

This story is perhaps amusing, but it mainly illustrates some fundamental differences between non-professional and professional adjudicators. A sense of what is fair and reasonable is of course a quality one would expect in any kind of adjudicator. In past days of imperialism and even today in the lowest level of the courts this

---

[1] In the following 'he', 'him', 'himself', etc. should not be understood as gender-specific, but also as including 'she', 'her', 'herself', etc.

may perhaps be sufficient qualification for a lay-judge. However, a professional judge has a far greater responsibility He must always state the reasons for his decisions, and both the decisions and their justification can be directly relevant and have impact on later decisions of other judges.

Concepts like 'fair and reasonable', not to mention 'just and equitable', are not always obvious for a case at hand, and even a professional judge may arrive at a wrong decision or give wrong reasons for a correct decision, though the legal systems of the world usually have options for appeal, etc. When the decision of a judge has to be *quantitative*, it is even more difficult to evaluate and compare to other decisions, and it sometimes happens that the public expresses its dissatisfaction with the quality of the sentencing of this judge or another.

In the domain of criminal sentencing it appears that even when giving more or less identical *reasons*, judges arrive at very different *conclusions*. In practice it appears that in many instances there is a great disparity in the decisions of sentencing judges – even in quite similar cases. It also happens that the very same judge will decide on vastly different sentences in similar cases occurring even over a relatively short time-span.

It is important to understand why this happens, and why some of those different conclusions cannot just be discounted as being erroneous. One of the reasons relates to different sentencing approaches. When passing sentence, a judge will usually take one or more of the following fundamental sentencing approaches into account: Retribution, Deterrence, Prevention and Rehabilitation (see Sections 2.1 and 2.3). Different judges may decide sentences in similar cases from different viewpoints and thus reach different results.

Sentencing policy is subject to changes. Usually such changes are initiated by legislators and the highest levels of the judiciary. But individual judges may also change their own, private sentencing policies with time. A judge we interviewed told us, that when he took up his position on the Bench, he often decided sentences with a view to rehabilitate the offender. "Now", he said, "I believe that the professional criminal cannot be rehabilitated, and it is only a matter of keeping him out of circulation".[2]

It is universally assumed that a judge passing sentence in a given case should follow trends established in sentencing of similar cases. There is therefore often some dissatisfaction with the sentencing process as it takes place today in the courts of many countries. A computer system for sentencing support could possibly enable the judiciary to pass sentences of greater uniformity, without, of course, impairing their freedom and independence.

---

[2] This private opinion also happens to reflect the general trend in many countries during the last 10–20 years. See also Lawton L.J. in Davies (1978) Cr. App. Rep. 207 C.A.

1.3. INTERACTION WITH ISRAELI JUDICIARY

We have been very fortunate that the Israeli judiciary at all levels has supported us in this research. Our work has not only been viewed with interest, but with active cooperation. The Israeli judicature has three levels, and several judges from the Tel Aviv Districts and Appeals Court (the intermediate level) have supplied us with expert knowledge about sentencing and tried our prototype systems.

Their main motivation is not their great interest in Computer Science in general or Artificial Intelligence in particular, but relates to their vehement opposition to the introduction of sentencing guidelines (see below, Section 3.1) in Israel. Some believe that any computer system we may come up with will be better than sentencing guidelines. Others fear that the politicians will attempt to make sentencing guidelines mandatory, but hope that the availability of a computer support system (any kind of system) will put an end to that attempt. The use of a computer support system would presumably always be optional and voluntary, so its introduction is not viewed as a problem.

1.4. CONTENTS OF PAPER

The outline of this paper is as follows. In section two we shall give some background relating to the criminology of sentencing. In section three we shall consider work that relates to non-intelligent computer systems for sentencing support. Section four will deal with intelligent sentencing systems, mainly rule-based. The next two sections describe the case-based approach and uses a system we have developed as an example. Section seven analyses and compares the various approaches with respect to maintenance, scale-up and efficiency.

2. Criminology of Sentencing

A sentence determines some form of punishment. It is therefore natural that we consider the the criminological aspects of punishment, especially as uniformity of sentencing (and hence of punishment) is a central issue in our work.

2.1. FRAMEWORKS FOR PUNISHMENT

Punishment and its purposes may be examined from three different aspects. The first one is the philosophy of punishment, the second is the framework of law and order in society, and the third is often called the sociology of punishment (Garland 1991).

The philosophy of punishments deals with questions like: What is the justification for inflicting punishment? What is a just punishment? etc. The second approach (law and order) deals with fighting crime. Within this framework criminal punishment has one specific aim: Reduction and prevention of crime. Punishment

is here considered, not from the viewpoint of justice, but from the viewpoint of efficiency. Which punishments will deter criminals? Can criminals be rehabilitated? etc. The third framework considers society and punishment, and the connection between them.

The four classical approaches to punishment, Retribution, Deterrence, Prevention and Rehabilitation form another type of classification of punishment commonly used by the judiciary and by criminologists:

> We have thought it necessary not only to analyse the facts, but to apply to those facts the classical principles of sentencing. Those classical principles are summed up in four words: retribution, deterrence, prevention and rehabilitation. Any Judge who comes to sentence ought always to have those four classical principles in mind and to apply them to the facts of the case to see which of them has the greatest importance in the case with which he is dealing (Lawton L.J., in: Sargeant (1974) 60 Cr. App. Rep. 74 C.A. at pp. 77–84).

These four sentencing aims are not easily incorporated within the first mentioned frameworks, as the concepts may conflict with each other: A retributive punishment may be just but not efficient, and vice versa. Some philosophers will argue that deterrence is not just. Why should one person be punished in order to deter others from committing crimes?

It should therefore be clear that different persons may hold very different views relating to the concept of punishments, and that society as such may change those views over time. As an example of totally conflicting views on punishment consider the following quotes from two eminent legal authorities.

> The punishment for grave crimes should adequately reflect the revulsion felt by the great majority of citizens for them: The ultimate justification of any punishment is not that it is a deterrent, but that it is the emphatic denunciation by the community of a crime (Denning 1953, para. 53)

> The idea that we may punish offenders against a moral code, not to prevent harm or suffering or even repetition of the offence but simply as a means of venting or emphatically expressing moral condemnation, is uncomfortably close to human sacrifice as an expression of religious worship (Hart 1963, 65–66)

It follows that when examining any concept relating to punishment, one should consider both the appropriate framework and the underlying criminological approach. Thus, the motivation for a given sentence in a particular case *must* be analysed with respect to these concepts, and the sentence cannot not be compared to the sentence of another – superficially similar – case without further inquiry. This now leads us to consider the problem of sentencing disparity.

2.2. UNIFORMITY OF PUNISHMENT

It seems obvious that the request for uniformity in punishment is more closely related to the search for justice than to prevention of future crime. The reaction to a serious crime ought to be stronger than to a petty crime, and people convicted of the same crime ought to receive the same punishment. One of the leading criminologists, Andrew von Hirsch, who is active in sentencing reform, has expressed this as follows:

> A sentencing system should seek to be just – or at least, to be as little unjust as possible. Claims about fairness, not. crime-control effectiveness, underlie the requirements of proportionality. (von Hirsch 1993, 103)

However, within a crime-fighting framework uniformity may be viewed as an obstacle. Should we wish to rehabilitate or deter offenders, then the punishment should depend more on the particular offender than on the particular crime he has committed. In many cases it is clear, that a prison sentence will only make an offender return to crime. In other cases it is equally clear, that only a prison sentence will prevent an offender from committing another crime, even if he has been found guilty of a minor crime. On the other hand, it has been argued that a non-uniform sentencing system can bring about feelings of frustration among prisoners, leading to unrest in the prisons.

The public at large expects some kind of fairness expressed through uniform sentences. However, the public also wants law and order, even though uniformity and effectiveness of punishment to some extent oppose each other. In reality the judges must seek some kind of equilibrium among the various approaches.

Such equilibrium is not easy to attain, and both politicians and the media see themselves involved. Israel is not the only country where the present judicial sentencing process is being criticised. In England, for example, the Home Secretary has strongly advocated mandatory minimum sentences for several crimes (Home Office White Paper, 1996), with equally strong opposition from the judiciary and other parties involved. Such a proposal is of course popular with a public that believes that its adoption will contribute to law and order.[3]

We shall give yet another example of the problems and the bias involved in finding the appropriate equilibrium. There is in Israel (and probably in many other countries) a general belief – strongly supported by specific examples in the media – that rape sentences are far too lenient. Information on offenders and sentences are stored in the central police files. No unauthorised personnel has access to this information, and authorisation is only given on an extremely limited basis. From our work in the sentencing domain we know for a fact that no statistical study has ever been made on the severity of rape sentences (or the lack thereof). Yet, the media make their assertions again and again based on a small number of cases they select for that particular purpose.

---

[3] The Home Office Recommendations have indeed been adopted by Parliament in the Crimes (Sentencing) Act 1997.

We must therefore amplify the views expressed in the introduction about the aims of a sentencing support system: The primary object of such a system should not be to help a judge attain some superficial measure of uniformity, but rather to consider uniformity in addition to the various sentencing policies explicitly formulated by legislators or otherwise.

Ashworth (1992) considers three techniques for reducing sentencing disparity, and is thus very relevant to our discussion. These techniques are: (i) Statutory sentencing principles, (ii) Judicial self-regulation, and (iii) Numerical guidelines.

(i) The first of these approaches is not relevant to Israel, but will be considered for the sake of completeness.

The idea behind this approach is to declare the leading principles of sentencing in legislation, but leave the judges to apply and to individuate these principles. Sweden and England are examples of countries in Europe, where this approach has been implemented. The penal code of Sweden identify Desert[4] as the primary basis of sentencing (von Hirsch 1992). A similar approach in England was formulated in (Home Office White Paper 1990)[5] from which we shall bring two excerpts:

> The first objective for all sentences is denunciation of and retribution for the crime. (Home Office White Paper 1990, section 2.9).

> Deterrence is a principle with much appeal ... But ... It is unrealistic to construct sentencing arrangements on the assumption that most offenders weigh up possibilities in advance and base their conduct on rational calculation. (Home Office White Paper 1990, section 2.8).

(ii) The second approach (Judicial self-regulation) is directly relevant to our discussion, as it forms the basis and justification for our case-based system. It will be considered in Section 6.2.2.

(iii) The third approach (Numerical guidelines) was briefly mentioned in Section 1.3. It is not relevant to Israel, but a computer system has been built implementing this approach in USA. It will be studied in the next section.

3. Non-Intelligent Systems

3.1. ASSYST

An extreme approach towards attaining uniformity in sentencing has been taken in the United States. The Sentencing Reform Act of 1984 is the statutory basis for the present Federal Sentencing Guidelines,[6] which became effective in 1987. These guidelines are very specific, defining ranges of sentences covering felonies and some misdemeanours. If a sentence passed fails outside the prescribed range the

---

[4] Desert is often used for Retribution. Roughly speaking Desert is Retribution without Vengeance.

[5] Later adopted by Parliament in the Criminal Justice Act 1991.

[6] There are significant differences among the federal guidelines and others adopted by many individual states. However, those differences need not concern us.

offended party may appeal, solely on these grounds. The United States Sentencing Commission, established in 1985, has the ongoing responsibility for the development and monitoring of the Federal Sentencing Guidelines, and recommends appropriate modifications to the United States Congress.

We shall not discuss the detailed reasons for the adoption of this approach, except that it relates to a general disappointment with the rehabilitation approach in force until then. When an American judge sentenced an offender to a prison term, he almost never had any idea of how long that offender would actually be behind bars (Reynolds and Beck 1993). It might not even be possible at the time of passing sentence to foresee how long that would be, as it depended upon the future behaviour of the offender. Under such circumstances the actual prison term of offenders who committed more or less the same crime could actually vary from one year to twenty. For early critique of the rehabilitation approach and of the sentence-guidelines approach now in force see (Martinson, 1974; Hudson, 1987). Later papers criticizing the guidelines approach are (Ashworth 1991; Tonry 1993).

Sentencing guidelines reduce the number of factors (and their combinations) taken into account. Obviously disparity in sentencing is reduced, for different courts will treat like cases alike. However, it may also be argued that the courts are forced to treat as alike cases that *ought* to be treated differently.

The approach to consistent sentencing as embodied in the sentencing guidelines has met with strong criticism from the American judiciary. It is felt that very little latitude is left for the judge in his decision, thus severely constraining any wish he may have to consider the particular circumstances of the case at hand. Judges have actually resigned rather than being forced to apply the restrictive guidelines (Freed 1992; Tonry 1993a).

A computer system called ASSYST (Simon and Gaes 1989) is based on these federal sentencing guidelines. It elicits from the user all the information required to make a decision, i.e. to determine a sentence according to the guidelines. It is a rather complex algorithm described in a general way in (Simon & Gaes, 1989). The authors of that paper mention that detractors have called the program a "Sentencing Guideline Calculator" (Simon & Gaes 1989, p. 199). That, however, should not be interpreted as a criticism of the program itself. ASSYST appears to be a professional, user-friendly and robust program for doing exactly what it is intended to do: Assist judges (or rather probation officers, as it appears) in applying the federal sentencing guidelines. The criticism is directed at the guidelines themselves, which make the sentencing process rather mechanistic.

The use of ASSYST in the federal courts has been reported as very successful. However, it is of no relevance in justice systems that do not subscribe to fixed sentencing guidelines.

3.2. LIST

This system was developed at the University of British Columbia (Hogarth, 1988). It takes a step in the direction of supplying a judge with relevant information, without actually prescribing a sentence. The database consists of sentencing decisions of the Provincial Court, Supreme Court and Court of Appeal of British Columbia. The user interface of the system uses menus extensively and operates in the following way:

(1) The user selects (from a menu) the appropriate legal domain (Criminal Code of Canada, Narcotics Control Act, etc.).
(2) He next selects the particular offence from a list of offences.
(3) He now defines a small number of offender characteristics. For robbery these would be: age-range, use of weapon (yes/no) and past record of violence (yes/no).

After thus having specified the facts of the case at hand the user may proceed in one of several ways.

 (i) He may request a histogram providing types of sentence ranges on the vertical axis for all cases in the database which match the facts of the case at hand. He may also request a table of the individual dispositions.
(ii) The user may view all or some of the relevant decisions of the British Columbia Court of Appeal. These cases are retrieved according to the type of the sentence.
(iii) The system permits the user to retrieve cases according to aggravating and mitigating factors relating to the offence, the record or the offender.
(iv) The system allows the retrieval of a summary of British Columbia Court of Appeal sentencing decisions which lay down general principles or specific factors which guide the exercise of discretion.

The statistical knowledge supplied by this system is said to be of great importance for a judge in order to determine what may be called an 'average' sentence with respect to the facts known to the system.

However, this approach has also met with some criticism. The statistical knowledge embodied in the Canadian system is based on a small number of characteristics as mentioned above. This does not suffice to express the actual complexity of the sentencing process. To quote a paper dealing with computer aids for sentencing:

> Without being too imaginative, the following dimensions could be used to define 'criminal record': (a) the number of previous convictions, (b) the recency of the last conviction, (c) whether the past record includes violent offences, (d) the length of time since the offender first was convicted, (e) whether the present offence was more serious than the most recent offence he had been sentenced for, and (f) the nature or severity of the offender's most recent sentence. It does not take a mathematical wizard to realise that if there are

even as few as three or four levels of each of these six variables, there are over 700 combinations of aspects of this one variable – criminal record. (Doob & Park 1987, p. 611)

It is well-known that there are many more variables like criminal record, each with several dimensions that in turn have different levels. The total number of combinations of all parameters would reach into the tens of thousands. A judge would first have to determine the values of all those parameters in the present case. This calls for much discretionary effort, and may easily discourage the judge. Next, statistics would have to be compiled for similar cases. But the number of previous cases corresponding exactly in all parameters to the case at hand would be so small, that the statistical data would not be significant.

Significant statistics could of course be compiled over a very long period. But this would be of little use: Sentencing data from, say, twenty years ago would probably not be relevant to the sentences being given today. Finally, detailed data relating to sentencing parameters is simply not available. We are not aware of any country where more than a few sentencing parameters are recorded and stored for future use.

Lovegrove is a criminologist who has worked extensively in the area of sentencing and sentencing models. He has argued that there are actually two problems with Hogarth's approach:

First it is acknowledged that there is disparity in sentencing, and it is important that a few disparate cases should not appear to be the norm. Second, because any classification scheme could not cover all the relevant case factors and their variations, the sentences for the cases falling within any one particular cell would cover the range. But where there were only a few cases representing a cell, one would have little idea where the cases lay in that range. (Lovegrove 1989, 42)

The question arises whether a judge passing sentence actually determines all those thousands of parameters for each case. It is obvious that the human brain works differently than a computer. It seems to be able to estimate the overall picture without a detailed analysis of each single parameter. This is the essence of human discretion, but a computer must take the detailed approach.

A different approach to a statistical sentencing system is proposed in (Doob & Park, 1987) depending on a much larger number of variables. According to (Lovegrove, 1989) this system nevertheless suffers from the same limitations as Hogarth's system. Furthermore, it calls for subjective estimates of several factors for determining the seriousness of the crime by the user, which is a "recipe for a system designed to foster and hide inconsistency of approach" (Lovegrove 1989, 47). Finally, why should a judge be bothered to apply discretion in order to determine various factors to be input to an advisory system? He might as well apply discretion directly to deciding a sentence.

The conclusion is, that the statistical data defining the starting point in the deliberations of a sentencing judge, while undoubtly superficially impressive, may simply be misleading. It is therefore quite interesting and somewhat surprising that other systems based on statistical information have since been built.

3.3. THE AUSTRALIAN (NEW SOUTH WALES) SYSTEM (SIS)

An Australian system: SIS (Sentencing Information System) has been built with the aim: "to promote sentencing consistency through the dissemination of information", and assuming that: "The provision of information on sentences imposed by the courts was seen to be a more acceptable method of achieving consistency than the establishment of a sentencing council or the development of sentencing guidelines". (Chan 1991, 139).

Values of the following variables must be input by the user:

1. Jurisdiction (local or higher court).
2. Act, section number or other specifications of the principal offence.
3. Whether other offences are taken into account in the sentence for principal offence.
4. Number of counts of principal offence.

Other offender characteristics may *optionally* be supplied by the user:

5. Prior criminal record.
6. Whether the offender was on ball or bond at the time of the offence.
7. Plea entered.
8. Age of offender.

The prior criminal record is chosen among four categories: (a) No prior convictions, (b) No prior convictions for an offence of the same type as the principal offence, (c) At least one prior conviction for an offence of the same type, (d) At least one prior conviction for an offence of the same type, which resulted in imprisonment. Type refers to one of the following categories: (1) Drug offences, (2) Sexual offences, (3) Fraud-dishonesty, (4) Property offences, (5) Offences against the person, (6) Driving offences, (7) Breach of order/escape.

The system supplies statistical data as well as other information similar to the output obtainable from the LIST system.

The builders of the system (Weatherburn et al., 1988) are well aware of Lovegrove's criticism of Hogarth's system quoted in the previous section, which they have attempted to answer:

> Since the statistics involve the *universe* of cases rather than a sample, the question of a small number of cases is immaterial. The absence or paucity of cases in a particular cell simply reflects the level of available precedents. (Chan 1991, 148 citing Weatherburn)

We have great difficulty accepting this argument. It is in any case not relevant whether a statistician would agree to it. The fact is, that judges are not experts on statistics and may misinterpret the data and give it far more weight than is scientifically correct. Many additional points of criticism of the system are found in (Greenleaf, 1991).

3.4. THE SCOTTISH SYSTEM (SIS)

A Scottish system also named SIS (Sentencing Information System) has been described in (Hutton et al., 1995). This paper makes no mention of the Canadian and Australian systems, nor does it relate to the criticism that has been voiced of the statistical approach. Its data-base contains all Scottish High Court convictions over a five year period (4500). The case-base thus relates mainly to first instance sentences, but the user may request information about possible appeal and its outcome. The case-base may also be restricted to appeals cases.

This is not a case-based system in the sense that it performs case-based reasoning (CBR). It retrieves cases from the case-base according to certain indexes and compiles sentencing statistics, using a relational data-base environment (SQL). For an overview of relational databases, SQL and their use in the legal domain, see (Zeleznikow & Hunter 1994, 29–38).

Initially the indexing addressed the important features in each offence category. However, in order to overcome the problem of insufficient data for statistical purposes convictions were reclassified using composite similarities (Hutton et al., 1995, 91). The latter were identified by discussion with the judiciary. Thus the problems formulated by Lovegrove and described in Section 3.2 also appear here.

The system operates in the following manner:

1. The user selects offence category and the classification of the offence within that category.
2. He may optionally choose offender characteristics and/or offence characteristics.

The distribution of sentences are then presented (as a histogram) on the basis of disposal: Imprisonment, probation, fine, etc. Each of 15 categories may then be further explored to show a distribution within that category, yielding details of sentence, reference to actual cases with low level details.

The prototype was used by four high court judges. They were especially interested in the potential of the system for informing sentencing decisions on unusual cases (Hutton et al. 1995, 93). It also appeared that the judges were particularly interested in displaying low-level details of cases, in order to examine their relevance to the case at hand. It is thus possible that the judges were alert to the danger of having a collection of 'extreme cases' in a small sample. A production system has now been completed, and is scheduled to be introduced in the Scottish courts within a short time (Wilson, 1996).

3.5. THE ISRAELI SYSTEM[7]

Despite the fact that the sentencing situation in Israel is fundamentally different from the situation in the USA (described above in Section 3.1), there are nevertheless people (mostly politicians) who believe that sentencing guidelines should also be introduced in the Israeli courts.

The previous president of the Supreme Court of Israel decided before his recent retirement, to initiate the development of a statistical information system by the Center for Computers and Law at Bar-Ilan University. This created a dilemma for us. On one hand the arguments given in the previous sections show that the use of such a system is scientifically suspect (to say the least), and our academic integrity should not allow us to participate in such a project. On the other hand, we did not want to jeopardise our good relationship with the Israeli judiciary, which has been very helpful in our CBR project described below.

After taking advice from criminologists, expert lawyers and judges, it was decided that an appropriate statistical sentencing information system could partially overcome the various defects described previously. This could be done, not necessarily by adding large amounts of detailed information about the case at hand, but by examining the case-base along one additional dimension only, the dimension of criminological sentencing approach.

The four classical principles play an important role in the study of sentencing. Thus (von Hirsch and Ashworth 1992), which is a respected collection of papers that consider the issue of uniformity and sentencing principles, is organized according to these principles. It is even more important that these principles also play a central role in the *sentencing process* itself. In chapter 2 we saw that legislation in some countries explicitly specify which of the four sentencing aims are to be considered central and which are less important. Furthermore, according to the saying by Lawton L.J. quoted in Section 2.1 every judge should consider these sentencing aims when passing sentence.

It is thus understandable that developers of the Israeli systems described below decided to incorporate these sentencing aims as basic parameters in their systems. However, there is absolutely no empirical data (or theoretical basis) for assuming that different sentencing approaches is the *only* or even the *major* reason for sentencing disparity. Ashworth (1995, pp. 60–80) discusses many other aims and principles that conceivably could be as important for a statistical analysis as the parameter based on the classical sentencing aims. In other words, adding this parameter is not necessarily an answer (even a partial one) to Lovegrove's criticism.

As in other systems the user must initially supply the following basic information about the case-base:

1. The relevant court(s)
2. The relevant paragraph(s) of the criminal code
3. The time period of retrieved cases (one or five years).

---

[7] It has not yet been named. By induction it should of course also be named SIS.

Information based on cases retrieved according to these parameters *only* is of extremely dubious value, as we saw above. Yet, the judiciary insists upon receiving it, and an existing prototype of the system will supply averages and histograms based only on the above information.

For a statistical sentencing base and given values of the above three parameters, the concept of an *average* can be defined in different ways:

1. The *mean* sentence is the sum of all retrieved sentences divided by their number (relates only to custodial sentences or other sentences with a numerical value).
2. The *median* sentence is the sentence dividing the set of retrieved sentences into two sets of equal size.
3. The *mode* is the sentence (or rather sentence-interval) appearing with the greatest frequency among the retrieved sentences.

Various kinds of statistical analyses may in principle be applied to the retrieved data. However, it has been argued by the judiciary that judges would not appreciate the results nor would they have time to examine them.

The user may request more precise information by checking one of the following items, hopefully leading to a reduction of the case-base which will yield more meaningful statistics:

Question: Sentencing approach was:

   Rehabilitation
   Deterrence of offender
   Deterrence of the public
   Isolation of offender
   Retribution

Again, appropriate averages and histograms will be presented to the user.

It was decided not to include case-descriptions at all. The relevant information needed to build a statistical database for the system may therefore be extracted from the central police files of offenders, which are continuously being updated. The actual computer program is therefore extremely simple (compared for example to the use of relational data-base techniques employed by the Scottish system), as no single case has to be retrieved.

Not enabling the user to access any descriptions of cases from the case-base appears to be a most serious defect. In the other systems the user can always check whether the facts of a particular case in the case-base are similar to the facts of the present case. Thus the Scottish judges who tested the prototype, actually looked at the individual cases. However, as one of our judges has pointed out, there may be a difference between a user's behaviour when experimenting with a system under test-conditions and the way he will use the system under realistic work conditions.

Judges may not find the time to read old records when faced with a stack of new cases to decide.

The police files do not at present contain any information on the sentencing approach. It was therefore decided that in future cases, judges will be asked to supply their answer to the above question, which will become part of the official protocol of the case. The scientifically dependable part of the program (if one can call it that) can therefore not become operational before sufficient statistical data has been collected.

It is intended that the database for the system will be built up from the present. No judge can of course be forced to answer the above question, but only those cases which do have answers will be selected and incorporated in the statistical database. Thus judges that choose not to answer these questions will know that they have no influence on future sentencing policy, and that their decisions will not be reflected in the statistical data available in future cases. It is hoped this will be sufficient incentive for the judges (or their clerks?) to supply answers to the relevant question in the appropriate part of the case-protocol.

Deplorably, it is thus known a priori that the system will have some serious defects:

(1) The lack of capability to check the actual facts of cases contributing to the statistics. (Judges may not find the time to use such an option, but it ought to be present).
(2) The lack of statistical confidence in the general results, retrieved without the sentencing approach parameter.
(3) The uncertainty whether the data relating to the sentencing approach parameter will ever be compiled.
(4) Despite these three defects one could perhaps hope that the use of the system by the judges would limit the disparity of their future sentencing. However, the present database reflects the current *non-uniform* sentencing situation, and its data is not meaningful. This is essentially Lovegrove's first argument in Section 3.2.

A prototype based on fictitious data has been built, and funds have been allocated for a real system. We hope that time and costs of development of the real system will be within the projected bounds. Collection of data relating to the sentencing approach parameter will depend on the cooperation of the judiciary.

3.6. CRITIQUE OF STATISTICAL INFORMATION SYSTEMS IN THE SENTENCING DOMAIN

Two such systems were developed (and abandoned) in Canada. Two other systems are actually functioning, and a third is under development. It is therefore important to summarize in a precise manner what the problems are with this kind of system.

1. Assume that the statistical data is retrieved based on only a small number of parameters. Then data for violent "professional" bank robbers would perhaps be considered together with the data for purse snatchers, giving a misleading picture of the fine-structure of the sentences.

2. Assume on the other hand that a very large number of parameters are used to retrieve the statistical data. The judge may not bother to input all those parameters in the first place. Even if he (or his clerk) does, the number of cases corresponding to given values of these parameters could be so small, that the retrieved data would not be statistically significant.

3. It is difficult to understand how uniformity can be obtained based on non-uniform data. A uniform database can be obtained by building it up from scratch using uniform data only. It is not easy to see how this can be done.

4. Judges are not professional statisticians and may attach too much weight to data than is scientifically correct. Users of a statistical computer system who are not expert statisticians or computer scientists, often trust facts and figures supplied by a computer, and automatically accept them as significant.

The principal source of guidance on sentencing (apart from legislation) consists of previous judicial decisions. Statistical systems seem to be a convenient way of summarising such material. As we have seen all such systems have serious defects, that cannot be overcome. They may actually supply misleading results, as users who are not expert statisticians or computer scientists, often trust facts and figures supplied by a computer, and automatically accept them as significant. However, such systems give decision support in a compact and readily surveyed form.

Additional material on statistical information systems is found in Section 7.

4. Intelligent Systems

4.1. RULE-BASED EXPERT SYSTEMS

Historically the first knowledge-based systems were rule-based, and it seems rather surprising that this technology was not applied to the problem of criminal sentencing long ago. In the general area of Al and Law many rule-based systems have of course been constructed. An overview of such systems is found in (Sergot, 1991), while (Susskind, 1987) gives a most thorough jurisprudential analysis of rule-based expert systems in law.

Apart from ASSYST, which could be rewritten as a rule-based system (but without heuristics), two systems are relevant, and will now be described.

4.1.1. *The System for Probation Officers*

A system has been built in Israel for the use of probation officers in recommending sentences for young criminals (Shapira, 1990). This system has been in existence and actual operation by the Youth Probation Services for several years. While technically not a rule-based system in the sense that it has a separate knowledge-

component in rule-form, it should nevertheless be classified as such. For the expert knowledge was originally formulated by the human experts in something close to rule-form. The sharp distinction between the knowledge-base and the engine of the system was lost during the translation to a programming language (DBase III).

The probation officer does not necessarily have to adopt the recommendation of the system, but she *must* give a detailed justification for any deviation from it. Of course, even if the probation officer adopts the recommendation of the system, the judge does not have to follow this recommendation, though she usually does (in 82% of the cases: (Shapira 1990, 161). The system has met with great success, it has been readily accepted by the probation officers, and is also said to save time and improve their efficiency.

Here we have an example of a knowledge-based system that actually assists in the sentencing process. It is, however important to consider the environment where this system operates. Probation officers are governmental employees who do not enjoy (and do not expect) the same kind of independence as judges. They must carry out orders and policy as formulated by their superiors with or without a computer system. If this policy happens to be expressed in rule-based form and included in a computer system, it makes their work easier, as experience indeed has shown. Furthermore, these probation officers are employed by the youth courts. Sentencing policy in those courts is much simpler than in courts for adults, and no conclusions about a similar system for adults can be drawn.

We mentioned above that the explicit form of the rule-base was lost in the actual implementation. It is usually considered a great advantage of rule-based expert systems, that the user may examine the applied rules, and thus convince himself if necessary that the answers supplied by the system are appropriate. This process is thus not possible in the probation system. We have interviewed several probation officers, who actually use the system, and asked whether they were bothered by this. From their answers it appears that, based on long experience, they implicitly trust the system and the answers it supplies.

4.1.2. *Bail Advisor*

Work is in progress in England on a rule-based system for bail-setting decisions (Hasset, 1993). It appears that decisions on whether to release a suspect on bail lack uniformity. While one magistrate (judge) will deny bail to an arrested person, another judge will set some minimal amount of bail for a suspect detained under almost the same circumstances.

On one hand the financial expense of remanding prisoners is very high, on the other hand there is of course great concern about persons who commit offences while free on bail. The criminological aspects of the project have also been described in (Burrows, Henderson & Morgan, 1994).

The English system under development uses the following approach. The magistrates are interviewed about the way they go about deciding about release on bail. This information is expressed in rule-form with the intention to build a rule-based

expert system. When a magistrate has to make a bail-release decision he would query the system, which would supply him with a definite answer to the question whether to release or not to release the suspect. Obviously the magistrate would not be forced to follow the decision of the system, but hopefully this decision would be an important guide for him, when handing down his decision.

The applicability of this extremely interesting system is at present inconclusive (Hasset, 1994). Also here it is important to consider the environment where the system is supposed to operate. Decisions on bail in the U.K. are made by the lowest level of judiciary, the magistrates. These are laymen of impeccable background, who have volunteered for this job. Thus, on one hand this group exhibits certain similarities with the probation officers, on the other hand the magistrates have many of the characteristics of professional judges.

4.1.3. *Other Work*

Berman & Hafner (1989) surveys the problems related to computer systems for sentencing. The paper also discusses ways of actually implementing rule-based expert systems for sentencing, and raises the possibility of using MYCIN-like weights. MYCIN (Davis et al., 1977) is one of the first medical expert systems developed. It associates a so-called certainty factor (Johnson & Keravnou, 1985) with each rule and combines these factors according to the laws of fuzzy logic. Berman & Hafner (1989) proposes to associate 'danger factors' to rules in such a way that a "jail term may be indicated when the danger factor exceeds a particular value" (p. 935).

The weighting approach of MYCIN and actually of most classical expert systems has not met with uniform approval (see Wellbank (1983)). Furthermore, experts in the legal domain do not usually reason in terms of numerical weighting schemes or probabilities (Ashley and Rissland 1988, p. 239), even though the final outcome of such reasoning in the case of sentencing can be numerical (e.g. length of prison sentence).

The JURICAS shell (van Noortwijk et al., 1991) has been used to create an advisory system in the area of remanding prisoners in custody.

4.1.4. *Critique of Rule-Based Systems in the Sentencing Domain*

The practical usability of rule-based computer systems in various domains has been under discussion for over twenty years, with opinions ranging from total rejection to total acceptance. We have surveyed two such systems above. The first of these has had a proven success, the second is under development.

Both these two systems deal with decision problems similar to criminal sentencing. However, the persons making those decisions are not professional judges. In the Israeli system they are social workers (youth probation officers), and must necessarily follow the regulations of their service. In the English case the lowest level of judiciary is made up of non-professional judges. They do not have to

follow any strict regulations imposed from above, and thus have the independence of judges. It is possible that they will agree to work with a rule-based computer system.

This, however, is not the case with professional judges. Our impression from speaking with judges is one of total rejection of the idea of using fixed rules for sentencing. The rules of a sentencing system would reflect the private view and opinion solely of the judge formulating the rules. These opinions may not necessarily be held by other judges. A panel of judges may not reach an agreement about sentencing rules, and even if they did, no judge would be obliged to follow those rules, or even to consider them while passing sentence. Adopting a set of rules by *legislation* is of course equivalent to the American approach through sentencing guidelines.

We may add some further reasons why the rule-based approach is not appropriate for sentencing advisory systems. These reasons relate to rule-based expert systems in general, and have been observed many times in the past (Wellbank, 1983).

 (i) Experts are often inconsistent in the sense that they do not practice what they say they do. In other words, even if they are willing to formulate rules, they do not always follow those rules themselves.

 (ii) Rules extracted from a panel of experts are often conflicting. Even a single expert often contradicts himself. In the case of judges these inconsistencies may reflect the actual inconsistent sentencing.

(iii) A prominent feature of rule-based systems is their capability of explaining the rests they supply when queried. However, the systems do not enable differentiation between rules of a technical nature and rules with conceptual expert knowledge, so the explanation of results becomes just an explanation of the formal deductions. Motivation for the application of specific rules is often lacking, as the systems do not have deep domain knowledge. (See, e.g. (Alty & Coombs 1984, 101–102; Davis & Lenat, 1982) on the expert system TEIRE-SIAS, or (Chandrasekaran & Mittal, 1982; Clancey, 1983)). We have seen for the probation officer system, that the capability of giving explanations is not critical. This cannot be true for judges. Should one really build a rule-based sentencing system, explanation of the conclusions must be perhaps *the* most prominent feature.

(iv) Extracted rules have semantic vagueness. It is not clear how they should be formalised in a computer program (Schild 1992, pp. 133–137). Whatever formalisation is decided upon, implies that discretion is ignored.

 (v) The explanation construction problem is computationally intractable. Building a completely new explanation from scratch is a very time-consuming process. Explanations are essentially causal chains. Small chains are reasonably easy to build, but the number of possible chains gets very large as the length of the chain goes up. For explanations the size that humans routinely create, the search space that a system that built such explanations from scratch would have

to serve is very large. If we had to build all (or even most) of our explanations from scratch, we would never have any time to do anything else (Schank, 1994). The case-based and model-based reasoning paradigms were developed as methods for bypassing reasoning from scratch. They will be considered in the next two sections.

4.2. CASE-BASED SYSTEMS

Case-Based Reasoning (CBR) is a problem solving approach by humans and computers where: "new problems are approached by remembering old similar ones and moving forward from there" (Kolodner 1993, xiv). Given a new problem, a CBR program (i) retrieves relevant cases from a case-base (using appropriately defined indices), (ii) chooses the most similar case, and (iii) adapts its solution to the new problem. The first step of CBR is often called Conceptual Retrieval.

There are some domains and tasks that are so well understood and well-defined that a small number of rules suffice to reason about them. Exhibiting the deduction process to the user would convince him about its validity, and some weighting scheme could also be applied. CBR would probably not be appropriate for such areas. It should be clear by now, that the sentencing domain does not satisfy these criteria, and CBR is therefore a good candidate for the underlying paradigm of a sentencing support system.

Several researchers have previously applied methods of CBR to the legal domain, but not to the area of sentencing (see, e.g., Rissland & Ashley (1987); Ashley (1990); Branting (1991); Skalak & Rissland (1992)). Ashley (1992) is a most comprehensive overview of the use of CBR in Law.

An early CBR program relating to sentencing was JUDGE (Bain, 1986). It used the sentencing domain to test cognitive theories of reminding and problem-solving. It was not intended as a program of practical use by judges. Our experience from interviewing judges closely corresponds to the cognitive observations in Bain's work (best described in Riesbeck and Schank (1989)):

These observations may be summarized as follows (Riesbeck and Schank 1989, pp. 98–110):

Prior record:

1. Judges try to predict the likelihood that an offender will repeat.
2. A judge will treat an offender more harshly if he believes that there is reason to predict recidivism.
3. Judges base their predictions about an offender's future behavior in part on the presence and severity of a prior record.

Forming explanations:

4. Judges seek reasonable explanations for criminal behavior.
5. Judges base their expectations about the arguments of the attorneys in a case on their own explanations of the offender's behavior.

The Victims:

6. A judge tries to punish an offender to an extent which is consistent in part with the amount of harm caused to the victims.

Experience-based generalizations:

7. The extent to which a judge considers certain features of cases and of offenders to be significant is a function of recent experiences he has in dealing with those features.
8. The extent to which a judge considers certain features of cases and of offenders to be significant is a function of prototypical experiences he has in dealing with those features.

Choosing a Structure for Sentencing:

9. Judges enumerate the set of choices which they believe they could impose as sentences.
10. The more a judge believes that an offender is likely to repeat a crime which involves victims, the more he will prefer to remove the offender from society by giving him a prison term.

Conditions on Choices:

11. Judges attach conditions to the sentencing choices which they enumerate.

Motivation of the Offender:

12. The process of explaining an offender's behavior involves inferring his motivations.

Case-based Reasoning:

13. Judges sometimes formulate sentences by using case-based reasoning.
14. When a judge applies a sentencing strategy from one case to another, he'll modify the sentence to accommodate feature differences between the cases.

Rationalization:

15. Judges stress the relative importance of specific features of cases to support the rationality of their sentences.

Indexing and Retrieval:

16. Judges retrieve cases from memory – their own cases as well as others – to compare to new cases.
17. Judges relate causal differences in cases to differences in the sentences which they give.

Some further conclusions about the behavior of judges when passing sentence are found in (Riesbeck and Schank 1989, pp. 111–113).

Further discussion of CBR-systems for sentencing, including the description of a system we have developed, is given in Sections 5 and 6, where we shall return to Bain's work.

4.3. MODEL-BASED SYSTEMS

The rule-based systems dealt with in Section 4.1 have a knowledge-base in rule-form, with rules that have been elicited from human experts. It is also possible to create a computer system, where the knowledge is based on a theoretical model. The theoretical knowledge may be expressed in rule-form or in any other convenient manner. It may be combined with other kinds of knowledge (e.g., a case-base), or appear in a stand-alone system. A general overview of model-based systems is Forbus (1988).

In the area of knowledge-based systems there are examples of the success of model-based ones, especially in combination with other kinds of knowledge-bases, e.g., case-bases (Kolodner, 1993). Criminologists actually build sentencing models (Wilkins et al. (1978); Fox (1981); and Lovegrove (1989)), though they may not agree among themselves about the validity of those models. So it should in principle be possible to build model-based computer systems for sentencing. DeMulder and Gubby (1983) describes a working system in the sentencing domain using a model-based approach (the Hulsman model), and an extremely interesting economic model of sentencing is described in Posner (1985). However, the judges do not believe in these models or trust them – as the criminologists are the first to acknowledge themselves. We shall therefore not discuss this kind of system any further.

We emphasise that our rejection of model-based systems in no way reflects on the *accuracy* of those models, which we have no intention – indeed no qualification – for analysing. In the introduction we stated that our object is to examine systems that judges would actually use. Model-based systems do not appear to fulfil that criterion.

4.4. OTHER PREVIOUS WORK (INTELLIGENT SYSTEMS)

A small number of other projects have dealt with computer systems and sentencing. We shall not consider these any further, as the systems discussed above are representative for our purposes. A system has been developed in Tennessee for sentence calculation (Reynolds & Beck, 1993). It assumes that the judge has passed sentence, and calculates the release dates for offenders. It illustrates how difficult it is indeed to compute the actual term of imprisonment in the US, as mentioned in Section 3.1 above. It is not of relevance to our problem.

5. Conceptual Case Retrieval

In the previous sections we have described the problems relating to sentencing advisory systems, and previous work relating to such systems, including a brief mentioning of case-based systems. In this and the following section we shall extend the discussion of the latter systems, including our own contributions.

5.1. BACKGROUND ON RETRIEVAL SYSTEMS

The first step of the case-based paradigm as considered in Section 4.2 is case retrieval. Textual information retrieval systems have been in existence and use in the legal domain for over twenty years. A user will state his requirement by specifying a word or a set of words connected by Boolean operators, and the system will retrieve those textual documents which satisfy the requirements. Two concepts are relevant to evaluating the retrieved results of a query (Salton, 1989)):

(1) Precision – the ratio of the number of retrieved relevant documents and the total number of retrieved documents.
(2) Recall – the ratio of the number of retrieved relevant documents and the total number of relevant documents.

Thus for a given query the precision and the recall should both ideally be one. That would signify that all relevant documents and only those have been retrieved. However, experience shows that the classical method of information retrieval can be problematic. A user is often interested in locating documents relevant to a certain *concept*. But texts and concepts may not have a one-to-one correspondence due to homography and synonymy, and the search operators for specifying relationships between *words* of text are crude approximations for specifying the relationships between the underlying *concepts* (Bing, 1987).

Instead researchers have developed conceptual retrieval systems where the search method is based on the meaning and significance of the retrieved documents: Conceptual retrieval is the process by which systems index and retrieve information using *conceptual structures* rather than text structures.

In the legal domain the term 'conceptual retrieval' has been used by several researchers, primarily by Hafner (see Hafner (1981) and Hafner (1987)). According to Hafner (1987) the knowledge-base of a conceptual retrieval system consists of three parts:

(1) A domain knowledge model, which defines the concepts the system should understand and know about, both legal and common-sense knowledge. Also relations among these items of knowledge are included in the domain knowledge.
(2) Individual case descriptors, which are descriptions of each case in the case-database structured according to the concepts of the domain knowledge model.
(3) A hierarchical rule system (called the issue/case discrimination tree in Hafner (1987)). This set of rules will enable the system to locate and retrieve the cases relevant to the problem at hand.

Given a conceptual retrieval system with three components appropriate to the sentencing domain and a sentencing case-base, a judge about to determine a sentence could retrieve *precisely* those cases relevant to a stated concept.

However, the creation of such a system is far from easy. Each of the three components specified above raises a different problem. With respect to the first

component: It is not possible to include all necessary legal knowledge and certainty not all common-sense knowledge in any present state-of-the-art system. As for the second component, the following view has been expressed by a leading authority:

> If intellectual indexing terms are assigned to documents as part of a related publishing scheme – as for WESTLAW and the Butterworths material in LEXIS – one would obviously include them. But if the costs are not met by such additional activities, one would hardly be able to justify the indexing costs by increased performance. (Bing 1989, 154)

Finally, concerning both the first and the third component, expert knowledge is necessary in order to create the appropriate hierarchical structures, but may not always be available.

Besides Hafner's work an attempt at formalising cases for the purpose of conceptual retrieval was carried out by Dick (1991). She applied Sowa's conceptual graphs in the area of contracts, and each case was analysed using the Toulmin argument model (Toulmin, 1958). Conceptual graphs have a well-defined semantics, but the work necessary to formalise any given case is considerable. Dick only carried out the analysis for a few cases. Mital et al. (1991) deals with conceptual information retrieval in litigation support systems. Turtle (1995) is a comprehensive overview of text retrieval, and Smith et al. (1995) is a recent paper describing a system involving conceptual retrieval and automatic indexing.

Statistical support systems could in principle be classified as conceptual retrieval systems, as the first step of their operation is a selection of cases according to some specified indices. They are knowledge-based systems in the sense that the indices used for retrieval are determined through knowledge of the sentencing domain.

5.2. OUR CONCEPTUAL RETRIEVAL SYSTEM

Keeping in mind the problems mentioned in the previous section, we have developed a computer system which uses conceptual retrieval in order to present relevant information from a sentencing case-base. There is no overlap or connection between this system and our statistical information system described in Section 3.5. The domain knowledge was elicited from an expert judge, the vice-President of the Tel-Aviv District and Appeals Court, who has many years experience on the Bench, especially in criminal cases. After many sessions and iterations the parameters that judges take into account when passing sentence, were determined. These factors were also found appropriate by our academic associates, a professor of law doing research in the area of criminal law and a panel of criminologists.

Eliciting information from experts is always a difficult enterprise. Our method was proposed by the judge himself. We decided, as it were, to create a questionnaire to be filled out by criminologists building a sentencing case-base based on actual trial protocols. In other words, the judge told us what a criminologist should look

for in the records, what issues and concepts relating to sentencing should be extracted and noted down. Obviously these concepts were the ones the judge himself considered relevant to the sentencing deliberations.

The sentencing parameters were arranged as nodes in a set of discrimination trees. One tree deals with features relating to the offender himself. In that tree, e.g., 'not-main-offender' is a father-node of 'weak personality', which in its turn is the father node of 'easily-influenced by others'. Other trees deal with the victim, the offence itself, mitigating and aggravating circumstances, and general issues. Other parameters relate to specific crimes.

It is obvious that the area of sentencing is associated with an enormous amount of both common-sense knowledge and domain knowledge. In fact, a complete model like the one proposed by Hafner [1987] would be almost impossible to create, even if we limit ourselves to only a few specific paragraphs of the criminal code. On the other hand, a hierarchical structure is both necessary and sufficient for a system to carry out an initial step of retrieval. We therefore decided to use the elicited domain knowledge only, without any additional common-sense knowledge. The nodes of the discrimination trees were taken as indices to cases in the case-base, i.e., these were the concepts a judge would be interested in.

We have limited ourselves to two serious crimes: Robbery and Rape. As indexing of cases is a technically big problem for large case-bases, we decided to establish a case-base spanning only the last five years. A larger time-span would raise problem of trends and changing attitudes in sentencing, which at this stage we do not intend to deal with. In Israel robbery and rape cases are heard in the District Courts, and may be appealed to the Supreme Court. The maximum sentence for each of these offences is twenty years imprisonment. This is one of the reasons for choosing these crimes, as sentencing disparity becomes especially pronounced when long custodial sentences are available.

For reasons to become apparent in the next section, we selected only Supreme Court cases, and only those where the appeal relates to the actual sentence. The prototype case-base covering the last five years contains less than a hundred cases, and manual indexing was thus possible. It was carried out by a qualified criminologist and a graduate computer science student, who had also been involved in the knowledge elicitation process.

The system assists the user to walk through the discrimination trees and check the nodes that are relevant to his case. The system then retrieves those cases from the case-base, which are indexed by the chosen nodes. Furthermore, retrieval is also carried out for nearest neighbours of the chosen nodes. Thus, if the user has checked the node 'not main offender' in the offender-tree, the system may also retrieve cases indexed by a sister-node, e.g., 'assisting after execution of crime'.

Obviously cases retrieved by neighbouring nodes have less bearing than cases which exactly fit the checked parameters. The system uses a Tversky-type similarity metric based on the contrast model (Tversky, 1977) to order the retrieved cases before presenting them to the user. For each case the system also presents

a list of the relevant parameters. The user can choose three levels of output: (1) a formulation of the ratio of the case, (2) the parts of the law-report relevant to the sentencing, (3) the entire law-report. Law-reports in Hebrew are not different from law-reports in English, and we shall not show any output from the system.

The expert knowledge of the system is contained in its discrimination trees, in its capability of retrieving not only directly relevant cases, but also cases indexed by neighbouring nodes, which may have some bearing upon the new case. The system was presented to several judges, who found it of far greater use than the classical text retrieval systems, but of limited application because of the case-base, which was too small and only contained appeal-cases. We observed that the judges recognized most of the retrieved cases when testing the system. This is not surprising, as the case-base consists of Supreme Court decisions, most of which would be landmark cases. We have already remarked in Section 3.5 that the readiness of the judges to read the retrieved law-reports or excerpts from these under test-conditions does not allow us to draw any conclusions about the applicability of the system under realistic work-conditions.

The system could have been further developed in several directions. Using Hafner's ideas much more intelligence could be added, and the case-base could be significantly enlarged. That, however, would necessitate the addition of vast quantities of domain-knowledge and especially of common-sense-knowledge. The present state-of-the-art of AI does not allow that. So we did not proceed with any further development of this system, and decided to use it not as a stand-alone system, but as a complement to the system described in the next section.

6. Case-Based Sentencing Systems

6.1. INTRODUCTION

In previous sections we have described various approaches to sentencing support systems. We shall now discuss the case-based approach in more detail and – as an example – describe a system of our own.[8] Despite some theoretical foundation for the CBR paradigm, systems that have been built over the years tend to be one-of-a-kind systems, with only a minimum of common features. It is therefore not surprising, that there should be several possibilities for implementing CBR-systems for sentencing support. Ours is definitely not the only possible method, and we shall therefore not dwell on technical details of the system in this paper. What we find of importance is to argue that the CBR paradigm seems to be the only really feasible one which may easily be accepted and actually used by the judges. Our system is an illustration only of how this may be done.

---

[8] This is the third Israeli system mentioned in this paper. The first one is the statistical information system described in Section 3.5 and the second one is the conceptual retrieval system described in Section 5.2. The second system was eventually incorporated in the third one.

We remind the reader that our desire is to build a computer system that may *support* a judge in deciding which sentence to hand down in a new case. Such a system should not supply its user with a single, definite answer (i.e., a proposed sentence). No judge would appreciate that. We propose an intelligent decision support system, where several approaches and ways of reasoning will be produced for the user, but the final choice will be his only.

6.2. JURISPRUDENTIAL FOUNDATION

Two interesting concepts from the domain of criminal sentencing are relevant to the building of a computerised sentencing support system. One of these is the concept of the 'tariff' as it appears in the writings and analyses of D.A. Thomas. The other relates to the English guideline sentences (as opposed to the American sentencing guidelines) handed down by the Court of Appeal.

6.2.1. *The Tariff*

It is traditional practice by the courts to establish a sort of scale – the tariff – which grades the severity of the penalty in accordance with what is regarded as as the gravity of the offence, without any individual consideration of the offender. Tariff sentences (if actually imposed) are fixed at points on this scale relative to the worst imaginable case of the offence in question, for which the maximum sentence is reserved.

The concept of a tariff is never given an exact numerical definition, though it is often used by writers on sentencing policy. Its meaning is best understood from the following quotes from the person who is considered one of the foremost authorities on this subject (Thomas, 1979; Thomas, 1982):

> A tariff sentence will be imposed when the sentencer wishes to emphasise to the public the gravity of the offence, while an individualised measure will be chosen where the object is to influence the future behaviour of the offender. (Thomas 1979, 14)

> The principles of the tariff constitute a framework by reference to which the sentencer can determine what factors in a particular case are relevant to his decision and what weight should be attached to each of them. Properly used they offer a basis for maintaining consistency in the sentencing of different offenders, while observing relevant distinctions, making appropriate allowances for individual factors and preserving adequate scope for the exercise of judicial discretion. (Thomas 1979, 29)

Thomas considers a number of offence categories. Within each category he assumes a variety of factual situations may occur. For each such situation he sets out the effects on the sentence. As we have concentrated on the offences of robbery and rape, we shall bring material relevant to these crimes exclusively. The section

on rape (Thomas 1979, 112–117) includes the following (all references to the large number of actual cases given by Thomas have been deleted):

> The scale of sentences extends from a usual minimum of two years to an upper limit of twelve years, although sentences both shorter' and longer are occasionally seen. The factors which advance the sentence within this scale include the degree of violence or threatened in the course of committing the act, the infliction of other forms of sexual abuse, the involvement of more than one defendant and the forcible abduction of the victim or the invasion of the victim's house. Conduct on the part of the victim which increases the risk of rape, such as willing participation in minor sexual activity or (to a lesser extent) imprudent behaviour such as accepting a lift in a car from strangers, is usually treated as a reason for some reduction in sentence. The age of the victim does not appear to have any substantial significance (except that the rape of a young child or an elderly woman may be an indication of the need for psychiatric investigation). The victim's previous sexual experience is not a major factor. The Court has said that 'it is not to be thought that whenever the victim of a rape is a woman of experience or of dubious moral standards it is any less an outrage to take that which the woman is not willing to give', although it is relevant to consider the extent of any physical or traumatic injury caused by the offence. The fact that the victim has forgiven the offender is not necessarily relevant. (Thomas 1979, 113)

One could possibly extract a set of rules for sentencing using Thomas' material, though quantitative information must be added. The general problems mentioned above in connection with rule-based systems would of course also apply here. In addition, rules based on Thomas' analysis could only be considered as secondary legal material.

But one can also imagine another kind of computer support system based on Thomas' analysis. Given the facts of a new case the system would retrieve an appropriate 'Thomas' (as it is often referred to in England). As Thomas' analysis is based on *actual cases* – which are all referenced in his writings, the next step is then to directly retrieve those cases that have bearing on the present case, perhaps even without going explicitly through a 'Thomas'. The contents of the 'Thomas' can function as *conceptual indices for retrieval* from the case-base. Some of those indices are of a special nature and relate to the particular type of crime being analysed. Others are general, and it is not surprising that we have discovered more than a partial overlap with the indices we have elicited from our expert.

There are, however, problems with this approach. As already mentioned, this kind of analysis obviously has no binding legal status (even though it references real cases), Thomas' work covers only some areas of criminal law, and it is directly relevant only to England. Additional problems will be discussed in the next section.

6.2.2. *Guideline Sentences*

Guideline sentences were pioneered by Lawton L.J. in the 1970s, and later on by the Lord Chief Justice, Lord Lane. Appellate decisions usually deal with the facts of a particular case only. But a guideline sentence sets out the parameters of offence for a particular category of crime, and then lays down a range of sentences for each such parameter. It does this, not by supplying a table of numbers, but in the textual style usually found in appellate judgments. The intention in delivering such guideline judgments may be deduced from the now famous saying:

> We are not aiming at uniformity of sentence; that would be impossible. We are aiming at uniformity of approach. (Lord Lane in: (1980) 2 Cr. App. R. (S) 177)

We shall give some examples of general guideline statements, from two appellate cases relating to rape:

> Rape is always a serious crime. Other than in wholly exceptional circumstances, it calls for an Immediate custodial sentence. . . . A custodial sentence is necessary for a variety of reasons. First of all to mark the gravity of the offence. Secondly to emphasise public disapproval. Thirdly to serve as a warning to others. Fourthly to punish the offender, and last, but by no means least, to protect women. The length of the sentence will depend on the circumstances. That is a trite observation, but these in cases of rape vary widely from case to case.

> Some of the features which may aggravate the crime are as follows. Where a gun or knife or some other weapon has been used to frighten or injure the victim. Where the victim sustains serious injury (whether that is mental or physical). Where violence is used over and above the violence necessarily involved in the act itself. Where there are threats of a brutal kind. Where the victim is very young or elderly. Where the offender is in a position of trust. Where the offender has intruded into a victim's home. Where the victim has been deprived of her liberty for a period of time. Where the rape – or succession of rapes – is carried out by a group of men. Where the offender has committed a series of rapes on different women, or indeed on the same woman. (Lord Lane in: Robert (1982) 4 Cr. App.R. (5) 8)

> For rape committed by an adult without any aggravating or mitigating features, a figure of five years should be taken as the starting points in a contested case. Where a rape is committed by two or more men acting together, or by a man who has broken into or otherwise gained access to a place where the victim is living, or by a person who is an a position of responsibility towards the victim, or by a person who abducts the victim and holds her captive, the starting point should be eight years (Lord Lane in: Billain (1986) 8 Cr. App. R. (5) 88)

As handed down by the Appeals Court, such statements were of course binding upon the lower courts. Given a particular case a judge could locate that case within the framework established by the guideline sentence. However, the judge would still have discretion to adapt the sentence to the special features of the case at hand.

There are several criminological and jurisprudential problems with this approach to sentencing. These problems will of course be directly relevant to any computer system seeking to incorporate the approach of guideline sentences.

(1) Only little guidance is given to the weighting of the aggravating and mitigating factors and their combination. That should not be surprising – such guidance cannot be given in an explicit form. That would reduce the guideline sentences to sentencing guidelines.
(2) The guidelines are not always numerically consistent (Ashworth 1984, 521–522).
(3) The guideline sentences only cover some areas of criminal law.
(4) Guideline sentences could perhaps reduce sentencing disparity in each separate area, but they do not form a general, overall sentencing policy.

The guideline sentences have been welcomed by the English judiciary, but the extent of their impact on sentencing disparity is not quite clear.

Our remarks concerning support systems using Thomas' analyses are also true with respect to guideline sentences (except of course that the latter are of a legally binding status). The guideline sentences may perhaps be used to extract rules for rule-based sentencing systems, but are far more suggestive of the CBR approach, and even explicitly define some of the indices for retrieval.

6.3. THE CBR PARADIGM AND SENTENCING

6.3.1. *General Overview*

Before describing some of the technical aspects of a CBR system for sentencing support let us attempt to outline the actual use of such a system, and in this way realize what components the system should be composed of and what capabilities it should include.

So let us imagine a judge about to pass sentence on an offender and wishing to use a CBR-system. In order to obtain some guidance he should be able to characterize the case at hand in some manner, so the system can retrieve a similar case. The system must of course convince the user that the retrieved case is indeed 'similar' in some definite sense. Using some kind of nearness-measure it should perhaps even indicate how near the retrieved case is to the case at hand.

The next step for the combined human-computer system is using the sentence in the retrieved case to arrive at a sentence for the case at hand. The fundamental assumption is that if the two cases are similar, the retrieved case will supply a guideline sentence in the sense of Lord Lane's sentences described in Section 6.2.2. The computer system should present an explicit comparison of the two cases in

a way that shows mitigating and aggravating aspects of the new case relative to the old case. The system can then indicate how to adapt the old sentence based on those aspects, and make appropriate recommendations. The actual adaptation will of course be done by the judge himself. The system should not recommend a specific sentence, although that is possible in principle. Some judges would not appreciate that.

Thus the first item to consider in order to create a CBR-system must be a collection of criminal cases from which to retrieve a relevant case. The following Section 6.4 will consider various possibilities for this case-base.

In order to enable case-retrieval the cases in the case-base must be indexed in some manner. The user of the system can then specify values of some of those indexes for the present case, and retrieve a similar case. Similarity means that the retrieved case has the same or 'near' values of the selected indexes. Thus the second point to consider is how to index the case-base. This will be done in Section 6.4.1.

It should be obvious that retrieving just a textual representation of an old case will not suffice. The comparison of the cases necessitates some structured representation of the cases. This structure is considered in Section 6.4.2, and retrieval is discussed in 6.4.3.

The actual reasoning, i.e., comparison of the cases, weighing the mitigating and aggravating factors and concluding how to adapt the old sentence to the present case is the final step, also briefly mentioned in Section 6.4.3.

6.3.2. *Our System: Short Example*

In order to illustrate the application of a CBR system for sentencing support we shall consider a concrete example executed on our (third) system. As the system interface and case-base are in Hebrew, there is no point in showing the actual input or output. Instead we shall show an edited English version in free format.

Consider the following story[9] (describing an actual case):

A young man had one night been out driving with a friend of his. The friend had then suggested they should rob a gas station. The man really had no intention of getting involved, but his friend eventually made him come along. They were later apprehended and found guilty. The man had unfolded the entire story in court, making a rather honest impression.

The system first elicits knowledge about this case by querying the user about the facts of the case. The word 'facts' is here used in the sense of conceptual indexes, i.e. nodes in the discrimination trees mentioned in Section 5.2 in connection with our (second) conceptual retrieval system.

The user must select one or more among the fundamental sentencing principles. This is one of the retrieval indexes in the case-base. For this case the judge indicated rehabilitation. As the system is interactive, a judge may experiment with several

---

[9] We shall use the word 'story' as the equivalent of the word 'case'. The motivation is (Schank, 1990).

sentencing approaches, view the suggested sentences of each approach and only then make up his mind and come to a decision.

Let us assume that for this robbery case the presence of the following facts (indexes) has been indicated by the user:

(1) use of weapon, (2) no criminal past, (3) additional offenders, (4) offender is young, (5) amount of money involved is small, (6) offender was influenced by others, (7) ...

The system will retrieve a case from the case-base relating to the robbery of a post office by two offenders. The user may retrieve a textual description of this case, should he wish so. However, the central output is a comparison of the two cases, i.e. a comparison of the conceptual indexes. This comparison consists of four parts:

A: Identical indexes: (1) use of weapon, (2) no criminal past, (3) additional offenders, (4) offender is young

B: Similar indexes (i.e. near neighbours in the discrimination tree): none

C: Indexes present in case at hand, but not in precedent: (1) offender was influenced by others (mitigation), (2) amount of money involved is small (mitigation), (3) offence committed at night (aggravation), (4) ...

D: Indexes present in precedent, but not in case at hand: (1) amount of money involved is large (mitigation[10]), (2) part of the money was retrieved (aggravation), (3) premeditation (mitigation), (4) ...

The system will state that the similarity between the cases is 'high'. Its final output is the recommendation that the sentence in the new case should be 'significantly lower' than in the retrieved precedent (five years).

The fuzzy values ('high', 'significantly lower', etc.) are precisely defined on appropriate scales. As explained previously, actual numerical values are not given unless specifically asked for, as some judges consider the main function of the system to indicate key characteristics (indexes) of the cases, and compare these. It is, however possible to request actual numerical results. In this example the system recommended a sentence of 1 year, while in the real case the judge decided on 13 months.

6.4. SYSTEM ARCHITECTURE

When building a CBR-system for sentencing there are in principle two possibilities for choosing a case-base. One can make it as large as possible, by including all

---

[10] Mitigation/aggravation in this part of the output means that the new case has a mitigating/aggravating factor when compared to the precedent.

available cases. This will facilitate the search for a case as close as possible (in some sense to be specified) to a new case, but does not necessarily alleviate sentencing disparity. For such a case-base would include cases from different judges, who execute their individual sentencing policies (and they are not even consistent with themselves). Given a new case, it may happen that the system retrieves one or more old cases with very non-characteristic sentences. This approach may even be worse than the statistical one criticised above.

On the other hand, one may build a case-base selectively, choosing only those cases whose sentences conform to the current sentencing policy. This would be in the spirit of Lord Lane's guideline sentences. When passing sentence a judge would thus not only retrieve past similar cases from his own memory, but the system would retrieve 'guideline cases' also to be considered. This raises the question who should make the choice of appropriate 'guideline cases'.

What we therefore propose is the establishment of a select sentencing case-base by a judicial commission, which has received appropriate authority through legislation. The case-base should be sufficiently large, so that for any new case arising it would be possible to find at least one 'near' case in the case-base. In practice this is of course impossible. There will always be new cases exhibiting characteristics never considered, or never considered together before. This implies that the case-base must be continuously added to. It must also be continuously culled, according to changing views of sentencing policies with time. All these activities should be the obligation of the judicial commission.

Even if all sentencing judges were to use a CBR support system with this authorised case-base, their sentences would nevertheless not be identical. However, two judges deciding two similar cases, and applying the same sentencing approach would receive identical recommendations from the system. More than that one cannot hope for.

When we decided to develop a CBR prototype for sentencing support, we had no case-base available, neither a large general one, nor a small select one. We therefore constructed the case-base from a collection of cases, which we obtained by interviewing judges. Our object was mainly to explore the way judges thought about old cases when passing sentence, and consider ways of automating reasoning using such cases, i.e. build an engine for a CBR system.

It was obvious that a system incorporating such a case-base could not alleviate sentencing disparity. However, the engine and the case-base are completely separated, and technically one can easily substitute another case-base (say a uniform one, in the same domain) for the present one.

The problems one has to deal with when building a CBR system are:

(1) Indexation: How may cases be retrieved by a case-based reasoner in order to give advice on quantitative results also for a new case?

(2) Knowledge Representation: How should criminal cases be represented in a computer.

(3) Reasoning: How should a retrieved case (or cases) be applied to a new case.

6.4.1. *Indexing*

The indexing problem raises the question which cases a judge is reminded of, when determining the sentence of a new case. A case, usually has one or more special and outstanding features. In the CBR community such a feature is called an 'anomaly' of the case (Kolodner 1993, 115), though its appearance may be quite ordinary. When encountering such an anomaly in a new case, a judge may be reminded about an old case with the same anomaly. Thus the anomalies could form the indices for a sentencing case-base.

In order to discover what the anomalies could be, we interviewed judges from the Tel-Aviv District and Appeals Court. We did not include the judge who had previously been involved in knowledge elicitation. The interview consisted in asking the judges to tell us about old cases of theirs, in the two areas of the criminal law we are concentrating on, Robbery and Rape. It then appeared that when a judge told us about a case, he would invariably come up with a heading for this case. He would supply that heading unasked, before or after telling about the case and sometimes in the middle. This heading turned out to be the most prominent anomaly of the case.

The case used as an example in the previous section was told by the judge presiding in the case. The judge explained his way of determining the sentence in this case, and ended by remarking: "That's what happens when you cannot say no".

It became apparent that the anomalies supplied by the judges were closely corresponding and sometimes even identical to the sentencing parameters (i.e., discrimination tree-nodes) described in the Section 5.2, thus confirming those structures. These parameters are also equivalent to the ones Thomas and Lord Lane have described. The judge's remark in the above story we construed to mean, that 'easily influenced by others' (a node in the offender's discrimination tree from the previous section) would be an index to the case.

In this case the index indicates mitigating circumstances. Our assumption is that when the judge mentioned above (and perhaps also other judges familiar with the story) encounters another case involving an offender with similar characteristics, he will be reminded of this case. He will also remember the severity of the sentence – or rather the *reasoning* behind that sentence as a factor (possibly among others), that may contribute to the decision in the case at hand.

We thus have sets of almost identical sentencing parameters obtained: (i) By elicitation from a senior judge. (ii) As headings supplied by other judges. (iii) In Thomas and Lord Lanes' sentences. (iv) By interviewing jurists and criminologists.

Kolodner discusses the indexing problem, i.e. how to retrieve the right cases at the right times. She defines two approaches for determining the retrieval indexes: The functional approach and the reminding approach (Kolodner 1993, p. 203). In the functional approach one considers the case-domain, the purpose of the retrieval and how the cases are applied. In the reminding approach one examines the way hu-

man experts are reminded of cases, in order to determine which kinds of descriptors are the important ones.

Our activities correspond precisely to these two approaches. Indexing knowledge elicited from the senior judge, extracted from Thomas and Lord Lane and obtained from expert jurists and criminologists is knowledge determined according to the functional approach. Indexing knowledge obtained through the interviews we conducted with the other experts is knowledge determined according to the reminding approach.

6.4.2. *Knowledge Representation*

Having determined the indices, the next problem is to create a good representation of a criminal case. There are several aspects to such a case: What happened, characteristics and motives of the persons involved, etc.

Consider first the narrative aspects. A 'story' is often considered as consisting of a sequence of *episodes*, i.e., events, actions, situations, etc., and the relationships among such episodes. It is dynamic with little or no hierarchical structure. Classical AI knowledge structures like semantic networks cannot in themselves suffice for representing stories. Such basic structures are appropriate for representing certain static aspects of the stories, but cannot cover the overall picture.

A *script* (Schank & Abelson, 1977) may actually be an appropriate form for knowledge representation for the legal process itself, as it may be used to describe ordinary and routine activities. It is also possible to use scripts to describe crimes, such as robbery or rape in a generic manner. However, scripts are not appropriate for describing the reasoning leading to the sentence imposed by a judge.

Also the Memory Organization Packet (MOP) (Schank, 1982; Kolodner, 1983), which generalises the script describes stereotypical events and does not enable the kind of explanations we seek.

Narrative understanding systems, e.g., CYRUS (Kolodner, 1981), BORIS (Lehnert et al., 1983; Dyer, 1983), and MEDIATOR (Kolodner et al., 1985) are not applicable here, as our object is not natural-language understanding or story-understanding, but the application of understood cases to a new case, which needs to be explained.

One could possibly use various kinds of logics, e.g., Episodic Logic (Schubert & Hwang, 1989). However, as the originators of such logics usually acknowledge themselves, much work remains to be done on these logics before they become applicable in practical systems.

Our decision was to consider the sentencing process as *case-based explanation*, and generalise the knowledge structure used for this approach.

6.4.2.1. *Explanation Patterns*. Swale (Kass et al., 1986a; Kass, 1986b) is a computer system which produces creative explanations for non-standard stories. Abe (Kass, 1989; Leake, 1992) is both a simplified and extended version of Swale. The system defines the concept of an 'explanation pattern' (XP) for a story. It uses the

explanation patterns for stories in the database to explain a 'gap' in the explanation of a new story. If these explanations cannot be applied in a straightforward manner, the system has a number of adaptation strategies.

As a concrete example, assume that the database contains the following two stories (originally given by Kass, and here considerably edited) and their XPs:

(1) A famous sportsman suddenly collapsed and died. The XP is: Unknown to everybody he had a weak heart.

(2) A otherwise healthy lady suddenly died. The XP is: Her husband killed her in order to obtain the insurance money.

Consider now a new story, based on an actual case: A famous racehorse (called Swale) suddenly collapsed and died. The 'gap' here is why the horse died suddenly. The system will adapt and apply the two previous XPs and suggest two possible explanations for the gap:

(1) Unknown to the owner and trainer the horse had a weak heart.

(2) The owner killed the horse in order to obtain the insurance money.

We shall call the sentence of a case the 'explanation' of the case. Determining the sentence of a new case means finding a similar old case and using its 'explanation' (adapting its sentence) to explain the new case.

The use of an XP in connection with a gap can actually work two ways. Given a story with a gap one can look for stories with an XP to explain the gap. Conversely, given a gap and its explanation one can look for an appropriate XP in order to *justify* this explanation.

A similar situation also occurs in the legal domain. Sometimes a judge will indeed decide on a sentence after considering the old cases. But sometimes he will decide on a sentence, and then see how to justify it (both to himself and to the world), by finding the appropriate precedents. This latter possibility is well-known and acknowledged by the judges themselves (they sometimes say they have a 'gut-feeling' of what the sentence should be). Obviously this justification is also an 'explanation' in the ordinary sense of this word.

6.4.2.2. *Judicial XPs*. An explanation pattern includes the following aspects (slots): (1) facts, (2) beliefs, (3) goal, (4) action (Schank, 1986; Schank et al., 1994). We have adopted this structure, and shall define it precisely below.

When a (decided) case is entered into the case-base, its XP is determined either by the justification given for written precedents, or by the explanations supplied by the judge, who told the 'story'. For a *new* case the facts and beliefs are supplied by the judge about to pass sentence in the case. He could also supply the purpose, but should have the capability of experimenting interactively with different criminological approaches. Thus the knowledge structure of both the case at hand and of cases in the case-base is the XP.

This approach is somewhat naive. Knowledge about explanation of simple cases may be represented by a single XP. However, legal cases are complex, and deciding a sentence necessitates the weighing of several factors. The concept of an XP appears to be insufficient for the kind of explanations we aim at creating. It cannot cope with the detail and complexity of most legal cases. We have therefore created a more comprehensive structure, which is made up of a basic MOP and a structure, called an MXP (Multiple XP), which consists of several XPs. We shall first show it for the above story.

basic facts MOP:

accused according to paragraph 402
found guilty
baseline sentence: 4 years
maximal sentence: 20 years
actual sentence: 1 year
appealed: no
purpose: rehabilitation

XP-1

facts: first offence
classification: not dangerous to public
action: reduce baseline sentence

XP-2:

facts: confessed
classification: seems trustworthy
purpose: retribution, rehabilitation
action: reduce baseline sentence

XP-3:

facts: easily influenced by others
classification: not dangerous to public
purpose: retribution, rehabilitation
action: reduce baseline sentence

A Multiple Explanation Pattern (MXP), is a comprehensive structure which is defined as a *collection of viewpoints* relating to the sentence in the same criminal case. Each such viewpoint relates to a fact that contributes to (increases or decreases) the sentence.

Each viewpoint is represented by an XP, and carries a weight relative to the other viewpoints. It is crucial to understand that an XP as such has no independent existence. Only the totality of the XPs forming the MXP describes the particular case and its sentence.

Each XP has the following slots:

1) Facts: This slot contains an index, which is a leaf in the index-hierarchy.
2) Classification: The values of this slot are (intermediate-level) indexes in the index-hierarchy.
3) Action: The value here indicates the action to be taken with respect to a baseline sentence. If, e.g., rehabilitation is the sentencing purpose, then the action-slot could indicate some degree of mitigation.

Two of the basic facts need some further explanation:

(1) Baseline Sentence: This is defined as 'the sentence given for this particular offence in the absence of any mitigating or aggravating factors'. This definition

is problematic, as different judges will choose different baselines. Indeed, this is probably one of the reasons for the existing sentencing disparity.

A judicial commission in Israel headed by a supreme court judge has recently completed its work and submitted a report to the Minister of Justice. In this report the majority opinion recommends the establishment of baseline sentences (according to the above definition) for all types of crimes in the interest of uniformity. Should this recommendation be adopted the creation of a uniform case-base will be easier.

When establishing the case-base the slots and basic facts of the MXP are filled out for old cases by a legal professional, preferably under the guidance of the sentencing judge himself. The sentence measured out in a case reflects the combinations of XPs in the MXP, though no explicit numerical formula for this combination is possible.

(2) Purpose: This indicates the sentencing principle applied by the judge (for each particular case in the case-base), or the sentencing principle the judge wish to apply (when using the system for a new case). Judges may of course use different approaches in different cases, and they only seldom state explicitly which sentencing principle they apply in a given case. This information must therefore be determined a posteriori by the experts analysing the case, perhaps with help from the sentencing judge.

In connection with the Israeli SIS (Section 3.5) we mentioned that Israeli judges will be asked to supply this piece of information, which will become part of the official protocol of the case. However, it is not clear whether the judges will actually comply with this request.

6.4.3. *Retrieval and Reasoning*

The retrieval may be implemented in different ways. One could, for example, consider the XPs of an MXP as similar to snippets (see (Kolodner 1988), (Redmond 1990), (Branting 1991)). However, our approach is to select a single, 'best' MXP using a Tversky contrast-measure, see (Tversky 1977).[11]

It is not our intention to give detailed descriptions and algorithms for the retrieval of old MXPs, choice of the 'best' one and its adaptation to fit the MXP of a new case, including a proposal of a sentence or a sentencing-range for the new case. In particular we have avoided any description and discussion of the determination of numerical weights that necessarily must be appended to the indexes. The concrete example given in Section 6.3.2 must suffice until detailed publication elsewhere. We are comparing the CBR paradigm with other paradigms in this paper, and indeed arguing for it, but not necessarily for the particular approach we have chosen.

---

[11] For a different approach to similarity of cases for sentencing purposes, see (Murback and Nonn 1993).

6.5. SUMMARY OF CBR APPROACH

We emphasise that our advocacy of this approach is based on the intuitive appeal this method will have to legal practitioners, and to the ease at which both engine and case-base may be constructed. We believe the small and select case-base approach is to be preferred, as it is easily implemented, but also the large case-base alternative can supply a judge with constructive advice when passing sentence, assuming of course that a large *uniform* case-base is available.

One point must still be clarified: Cases in a case-base may appear to be conflicting, just as rules may conflict. There is however a major difference between these two kinds of inconsistencies. Cases in the case-base reflect real legal situations, and legal experts are accustomed to resolve conflicts in case-law. They also deal with inconsistencies in statutory law, but the sentencing rules in a rule-based expert system are not the law, they have no legal standing. They are superficial creations, that carry no more weight than the individual judge decides to assign them (which may be none at all).

It is thus possible and quite acceptable, that the system will create one line of sentencing recommendation based on one retrieved case, while another retrieved case will yield another recommendation. The judge will make his own decision accordingly – this is the essence of human discretion, which no system has.

7. Comparison of Approaches

At the end of each section dealing with a particular kind of support system, we have summarised our criticism of that particular kind of system. We shall now discuss some additional features of sentencing support systems, and summarise our findings. We shall omit any considerations of the American sentencing guidelines program, as it is of no relevance to our main concerns. Nor shall we deal with model-based systems, for reasons given in Section 5.3. Thus we shall consider three kinds of systems:

 (i) Statistical Information Systems
(ii) Rule-Based Systems
(iii) Case-Based Systems

We shall consider various aspects of those systems. The financial and budgetary ones relating to the establishment and maintenance of the systems are of central importance, but will be considered only in passing. The aspects we shall examine are:

(1) Maintenance and Scale-up
(2) Complexity and Efficiency
(3) Acceptance by Judiciary

7.1. MAINTENANCE AND SCALE-UP

These two issues are often neglected in AI work, giving rise to some of the strongest criticism voiced against AI. Often a prototype system is constructed, but no consideration is given to the feasibility of updating and maintaining its knowledge-base from time to time. It also happens, that a prototype system is implemented with a knowledge-base appropriate for a 'toy-world', but the problems arising in an attempt to apply the system to the real world (scale-up) are not dealt with.

Concerning the transition from a prototype to a real world system there are two problems to consider.

(A) Is it feasible to obtain all necessary knowledge.
(B) Can the system at all function with a real world knowledge-base.

With respect to maintaining a support system, there are actually three separate problems to consider.

(C) Who decides when to update the system, and what knowledge to add, delete or exchange.
(D) How soon will the changes be reflected in the advice given by the system.
(E) How easy is it to carry out such changes.

Only very recently has consideration been given to some of these problems in the area of AI and Law (Bratley et al., 1991; Berman & Hafner, 1995; Rissland & Friedman, 1995). We shall deal with these aspects for all kinds of systems mentioned in this paper, notwithstanding our preference for a certain kind of system.

7.1.1. *Statistical Information Systems*

(A) Scale-up entails the collection of great numbers of cases from the relevant jurisdictions. Today this material would be available on magnetic media in many countries. If case retrieval is to be enabled, much work must be invested in indexing and other technical preparations. This is a problem of availability of expert manpower and sufficient funding.

(B) Problems of scale-up are heavily dependent upon the system architecture. While the Israeli system is very simple, the approach using existing relational database environments can lead to unwieldy procedures for maintenance. Nevertheless, even for very large case-bases present-day technologies should enable the straightforward establishment (and maintenance) of database systems.

The editorial in Crim, L. R., March 1997 directs the attention to another problem:

> The High Court of Justiciary in Scotland has given relatively few authoritative judgements on sentencing, but in other countries, e.g. in England, the Court of Appeal's decisions are manifold and ought to be followed. However, some

judges maintain that there is a big difference between the sentence levels up-
held by the Court of Appeal for certain crimes and sentence levels in practice.
(Crim, L.R. 1997, pp. 153–154).

The New South Wales system considered in Section 3.3 keeps separate data-
bases for sentencing practice and appellate judgements, and each country wishing
to create a statistical sentencing database must decide whether this is indeed the
architecture which is appropriate for the local situation.

(C) Presumably there would be no need for any prior examination of cases to
be added to the system, as the more cases included the better. Thus the process of
maintenance is an automatically on-going process, dependent only on appropriate
funding.

(D) The major problem with statistical systems (in addition to the ones already
detailed in Sections 3.2–3.6) relates to the extent to which the system reflects cur-
rent sentencing policies. The statistical analysis necessarily draws data not only
from very recent decisions, but also from old cases still present in the case-base.
Such cases should be discarded when their sentences are no longer considered
relevant, and somebody has to make that decision. Finally, if too many old cases
are deleted from the case-base, the remainder may no longer form a statistically
significant set.

(E) Much work is needed to review old cases, and it is not clear who should be
responsible for such work.

7.1.2. *Rule-Based Systems*

(A) Scale-up of a rule-based system entails establishing rule-bases for all types
of offence. Even if one were to imagine that agreement can be reached among
the judges about the formulation of such rules with respect to a single type of
offence, one can hardly expect agreement about general rules relating to the entire
sentencing domain.

(B) Assume that agreement can be reached, and that the rule-base is properly
constructed. This means that the rule-based system exhibits a strict division be-
tween the engine and the knowledge-base. Not all rule-based systems observe this
feature, e.g. the Israeli system for probation officers described in Section 4.1.1.
In the former case, scale-up will be additive (and hence relatively easy) in the
following sense: A rule-base for the entire criminal domain will consist of a general
part and the union of smaller rule-bases relating to each individual type of crime.
Thus going from a system addressing only selected criminal offences to a larger
system addressing additional offences, will imply only the addition of rules, and
no change in already existing rules. With proper annotation of the rules such a
system would not exhibit special technical problems. In the latter case, scale-up as
such may be impossible.

(C) We have previously (Section 4.1.4) expressed the view that no initial rule-
base can be constructed, which all the judges would agree upon. Such disagreement

would of course also carry over to the problem of maintenance: Who shall decide which rules to add, delete or change.

(D) If agreement on maintenance is reached, changes in the knowledge-base would of course have immediate effect on the advice of the system.

(E) If a division exists between-engine and rule-base and the rules are properly annotated, maintenance can be carried out. If not, updating the rules may be a technically difficult problem. For a very large system it may even be an impossible task. There is evidence that rule-based systems with more than a few thousand rules cannot be properly maintained.

7.1.3. *Case-Based Systems*

In Section 6.4 we distinguished between two types of CBR systems for sentencing. One of those would have as large a case-base as possible, while the other one would consist of a much smaller collection of select landmark cases. The points we discuss must be considered separately for each of those possibilities.

(i) Large Case-Base

(A) As we have explained above for statistical information systems, the collection and preparation of cases is a technical and financial problem. Indexing is a problem very relevant to the scale-up from a small prototype of a CBR system to a system with a large case-base. Automatic indexing is not feasible at present, though much research is being done on that topic using methods of Natural Language Processing or other statistical approaches (Chen, 1993; Ginsberg, 1993; Schweighofer, 1995). We are therefore dealing with manual work, which has to be carried out by experts.

(B) Whether a case-based reasoner would be able to function with a very large case-base remains to be seen. Most CBR systems built so far have been prototypes only, and it is difficult to foresee which problems may arise if the case-base includes tens of thousands of cases.

(C) The extremely big problem we described for statistical information systems also exists here. Who shall decide which cases to delete, and who will have the expertise, time, etc. to carry out a constant review of a large case-base.

(D) Changes in sentencing policy will be reflected the moment new up-to-date cases are added, and old, no longer relevant cases are deleted.

(E) The actual changes are at most a technical problem.However, this problem may be severe, if nodes of the index-trees have to be changed. Such a change will imply a similar change in the indexing structure of the individual cases of the case-base. This is no easy task, which at present cannot be carried out by an automatic procedure.

Changes in the index-tree structure would not be a frequent occurrence. However, changes in the weighing function (recalibration) could sometimes be necessary. This problem is technical, and not too severe.

(ii) Select Case-Base

There are two possibilities relating to which cases should be included in a select case-base for sentencing.

 (1) A judge may be willing to include only his own previous sentencing decisions in what will then become his own private case-base. It would not contribute to the aim of attaining general uniformity in sentencing, but at least the judge's own sentences may become uniform.
 (2) The establishment of a public panel, perhaps consisting of judges, practising lawyers, academics and legislators. This panel would decide which new cases to add and which old cases to delete from the case-base, thus ensuring that the case-base will reflect the sentencing policies in a faithful manner. It is an open question whether such a panel could reach an agreement about the inclusion of cases.
 (A) Without any experience at present, it is impossible to say how a public committee would reach agreement on which cases to include in the select case-base. Obviously a lot of politics would be involved. Technically, however, there are no problems.
 (B) CBR systems exist which function well with small case-bases.
 (C) As already explained a public committee will make the decisions. The actual update is technically straightforward
 (D) Changes in sentencing policy will be felt immediately with the appropriate changes in the case-base.
 (E) Updates are easily carried out. Even changes in the index-tree structure and appropriate changes in the indexing of the cases in the case-base would not be extensive, as the case-base is assumed small.

7.2. COMPUTATIONAL COMPLEXITY AND EFFICIENCY

7.2.1. *Classification of Problems*

The efficiency of a sentencing support system may be considered from three points of view.

(A) The effect of a system on the expenses of the justice system. It is a well-known feature of the courts of law in many countries that they operate at great cost and usually with a great backlog of cases. It is possible in principle that the introduction of computer systems of the type described above may reduce those costs.

Consider for example an expert system, which identifies and recommends bail decisions for individual cases. A prototype for such a system is described in (Hassett, 1994), and was discussed in Section 4.1.2. The financial burden imposed on the state by pre-trial detention is considerable. A computer system could perhaps identify the persons not to be detained in a more efficient manner than today, and thus lower the costs

However, it is far from clear whether this indeed will be the case. Experience shows that the introduction of computer systems usually increases the amount of information available to human users of the system, but does not decrease the costs. There are also many examples of environments where throughput was not increased through computerisation.

We therefore do not believe any conclusions can be drawn about the three types of sentencing support systems with respect to costs and expenses, especially as the costs of establishing and maintaining the system also have to be taken into account.

(B) What is the utility of the system? i.e., does use of the system actually improve the work of the judge using it? Does it raise the quality of his decisions? To establish criteria and benchmarks for the utility of intelligent computer systems in the legal domain is a general problem, that has not been dealt with so far. At present the only possible method to examine this issue must be by interviewing the users. No such data is available for any system mentioned in this paper.

(C) The purely computational aspect, i.e., whether the system interacts sufficiently fast with the user. As no data is available concerning the first two issues, we shall only deal with the third one for the three kinds of system we discuss.

7.2.2. *Comparison of Systems*

Statistical Information Systems:

In the case of the Scottish prototype the developers admit to severe limitations of resources [Hutton et al. 1995, 94], even though the case-base is quite small (4500 cases). We are concerned this will always be the case, when database technology is applied to what is essentially a straight-forward statistical problem. One of the main justifications of the Israeli approach to system architecture is the capability of attaining computational efficiency even with a large sentencing base.

Rule-Based Systems:

The efficiency of even very large rule-based systems has been well established. There is no reason why a rule-based system in the sentencing domain should not be efficient, once the rule-base has been created, especially if the rule-base is partitioned according to the crimes.

Case-Based Systems:

A CBR system with a small and select case-base will clearly be fast, even if its engine is not computationally efficient. For systems with a large case-base no previous experience is available.

7.3. ACCEPTANCE BY THE JUDICIARY

We believe that the major principle of guidance in constructing a sentencing support system should be its ultimate acceptance and actual use by the judges. A prototype will of course always be of theoretical interest. But there would be no

real purpose in fully developing such a computer system, if it would be applied by just a very small number of judges, used only by researchers, or not used at all.

It may not be easy to gain general acceptance for a decision support system by the judiciary, though the use of computers is a common phenomenon in a court today. The principal use of computers is for word-processing and sometimes for textual information retrieval, using a system like LEXIS or Westlaw. The acquisition of an expert advisory system should therefore not be problematic, for it does not introduce completely new technologies.

However, a system for sentencing support may be met with (perhaps passive) resistance for quite different reasons. One of the judges we interviewed in the project described above, told how both supplying and receiving sentencing information have been met with opposition by many judges. As president of the Tel-Aviv District and Appeals Court she had requested the judges of the court to send her short monthly reports about their criminal cases, stating briefly the facts and the sentences. These reports were to be distributed among all the judges of the court, with the intention of gaining if not uniformity, then at least a common and up-to-date basis for sentencing. Only few judges complied with her request, several told her they were against *receiving* (not to mention supplying) such information, and after some months she decided to stop the experiment.

When asked the reasons for the refusal of the judges, she indicated that perhaps these judges did not believe brief summaries were of any use, and that nobody had the time to write (or read) long protocols of other judges.

There appears to be a basic inconsistency here. We have already pointed out (in Section 2.1) that sentences should be analysed with respect to the motivation of the judge. Each individual judge is influenced and reacts differently to the impression made by the offender, the witnesses, etc. in court. Such influences are often very subtle, and can never be apparent from the records. Even the inclusion of a sentencing approach parameter is only a crude summary of this non-explicit information.This is of course well-known to the judges themselves, and may be the reason for the rejection by some of them of only short summaries. Such judges might presumably not want to know what a computer system has to say about the sentencing possibilities in a given case.

On the other hand, *all* the judges we interviewed agreed that statistical sentencing data is 'a good thing' and should be made available, even if based on only a small number of parameters. The initiative to add information on sentencing approach comes from our criminologists, and lack of this parameter did not seem to bother the judges.

It should be obvious that no judge can ever be *forced* to use such a system, and more subtle ways would be needed in order to convince the judiciary to use a decision support system. At this point let us just remark, that as the acceptance of this kind of system is problematic for external (i.e. criminological and jurisprudential) reasons, a software developer must take good care that it should at least be

extremely user-friendly and supply the user with sufficient information so that he should not also have to apply manual labour for searching precedents, etc.

7.3.1. *Statistical Information Systems*

We have already explained above that the judges seem to want precisely this kind of system, despite the lack of confidence in the data. The problem here seems to be that this kind of system could be too warmly accepted, in the sense that too much trust could be uncritically placed in its data.

7.3.2. *Rule-Based Systems*

It is difficult to imagine that rule-based systems can be accepted by the professional judiciary. There is of course always a possibility to impose a set of rules by legislation. However, the system thus obtained would be equivalent to a sentence guidelines system (Section 3.1), an approach not relevant to our present considerations.

7.3.3. *Case-Based Systems*

When approaching the problem of building a case-based sentencing support system in Israel or in most other countries outside England, one has neither Thomas nor guideline sentences to build on. There are, however, a large number of reasons for using the CBR approach. Much of what is usually said in favour of this paradigm does not amount to more than hype, in the sense that it may seem intuitively correct, but has not been backed up by real proof. Nevertheless, in the legal domain it is particularly convincing, as the use of previous cases is a central aspect of legal reasoning in general, and in the area of sentencing in particular. We shall therefore make the following observations:

(1) Judges are accustomed to work with cases, to apply them and to distinguish them. A case-based advisory system presents the judges with real cases and sentences, not made-up rules. If judges are at all ready to use a computer for sentencing support, cases is the natural medium for conveying information.

The reason for this is that cases represent experienced situations. When a similar situation arises, those decisions and the knowledge that went into making them provide a starting point for solving the problem the new case poses. In other words, using the CBR approach it will be possible for a judge to determine a sentence based on general standards but also to consider the individual circumstances of the offender and the case at hand.

(2) When a judge has to pass sentence in a new case, he often considers old and similar cases and may adapt an old sentence to fit the new case. These old cases are not only the (perhaps binding) precedents he must consider, but primarily cases from his own experience, or cases his colleagues have told him about.

A judge we interviewed told us how he actually keeps a card-index of his old cases and their sentences, which he flips through when he has to pass sentence in a new case.

Judges may not always be reminded of the most appropriate set of cases when they reason. There is evidence, that when there is much to remember, people cannot always access the right cases when they need them (Brooks et al., 1989; Gentner, 1989). A CBR system for sentencing will make the appropriate cases available, thus acting as an auxiliary memory for the judge. It will also help toward an analysis, but of course leave the final decision to the user.

Using a CBR system with a case-base made up of the judge's own cases would not be so very different from using a card-index. Of course, the only kind of consistency and uniformity to be obtained would be of the judge with himself. However, a system with this kind of private case-base could be a transitional step. It would accustom the judiciary to use a CBR system, where the case-base eventually would be made up by a public committee.

8. Conclusion

We have presented and analysed some approaches to advisory computer systems for criminal sentencing. From the jurisprudential and criminological background it should be apparent that there are many problems that arise when attempting to create such a system, but no ideal solution. In fact, most problems are not computational and do not arise from the limitations of computers or even from the area of Artificial Intelligence. They relate to inherent problems of the domain.

It is our opinion that among these approaches the case-based one is the most practical and feasible one. It is not perfect, but it seems optimal. This takes into account the objective problems of other kinds of systems, and the opposition from the judiciary itself to these other kinds of Systems. A CBR system is a tool which will not be too unfamiliar to legal experts, who anyway work with cases, and it is relatively easy to build.

Two major questions arise with this kind of system:

1) Can agreement be reached about which cases to include in the case-base?

2) Will the judges be ready to devote sufficient time to read and consider the analysis and arguments of the system?

It is our hope that positive answers may be found to both questions. The introduction of our present prototype on a limited trial basis in the Israeli courts will test this hypothesis.

Acknowledgements

References

Alty,. J.L. and Coombs, M.J. 1984. *Expert systems*. Manchester: NCC Publications.

Ashley, K.D. 1990. *Modelling Legal Argument*. Cambridge: MIT Press.

Ashley, K.D. 1992. Case-based reasoning and its implications for legal expert systems, *Artificial Intelligence and Law* 1, 113–208.

Ashworth, A. 1984. Techniques of guidance on sentencing, *The Criminal Law Review* 519–530.

Ashworth, A. 1991. Sentence reform structures, In M. Tonry (ed.), *Crime and Justice – A Review of Research*. Chicago: Chicago University Press, 16: 181–241.

Ashworth, A. 1992. Three techniques for reducing sentence disparity. In A. von Hirsch and A. Ashworth (eds.), *Principled sentencing*. Edinburgh: Edinburgh University Press, pp. 282–290.

Ashworth, A. 1995. *Sentencing and Criminal Justice*. 2nd edn., London: Butterworth.

Bain, W.M. 1986. *Case-based reasoning: A computer model of subjective assessment*. Ph.D. thesis. Yale University.

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

Berman, D.H. and Hafner, C.D. 1995. Understanding precedents in a temporal context of evolving legal doctrine. In *Fifth International Conference on Artificial Intelligence and Law (ICAIL-95)*. College Park, MD: ACM Press, pp. 42–51.

Bing, J. 1987. Designing text retrieval systems for conceptual searching. In *Proceedings of the First International Conference on Artficial Intelligence and Law (ICAIL-87)*. Boston: ACM Press, pp. 43–51.

Bing, J. 1989. The law of the books and the law of the files. In G.P.V. Vandenberghe (ed.), *Advanced topics of law and information technology*. Amsterdam: KIuwer, pp. 151–182.

Branting, L.K. 1991. Reasoning with portions of precedents. In *Proceedings of the Third International Conference on Artificial Intelligence and Law (ICAIL-91)*. Oxford: ACM Press, pp. 145–154.

Bratley, P., Fremont, J., Mackaay, E., and Poulin, D. 1991. Coping with change. In *Proceedings of the Third International Conference on Artificial Intelligence and Law (ICAIL-91)*. Oxford: ACM Press, pp. 69–76.

Brooks, L.R., Allen, S.W., and Norman, G. 1989. The multiple and variable availability of familiar cases. In C. Hammond (ed.), *Proceedings: Workshop on case-based reasoning (DARPA)*. San Mateo, California: Morgan Kaufmann, pp. 130–132.

Burrows, J., Henderson, P., and Morgan, P. 1994. Methodology: A project to improve bail decisions. Presented at: *Conference of the American Society of Criminology*. Miami, USA.

Chan, J. 1991. A computerized sentencing information system for New South Wales courts, *Computer Law and Practice* 137–150.

Chandrasekaran, B. and Mittal, S. 1982. Deep versus compiled knowledge approaches to diagnostic problem-solving, *Proceedings AAAI-82*. Los Altos, California: Morgan Kaufman Publishers, pp. 349–354.

Chen, H., Lynch, K.J., Basu, K., and Ng, T.D. 1993. Generating, integrating and activating thesauri for concept-based document retrieval, *IEEE Expert* 8(2), 25–34.

Clancey, W.J. 1983. The epistemology of a rule-based expert system – a framework for explanation, *Artificial Intelligence* 20: 215–251.

Davis, R., Buchanan, B., and Shortliffe, E. 1977. Production rules as a representation for a knowledge-based consultation program. *Artificial Intelligence* 8, 15–45.

Davis, R. and Lenat, D.B. 1982. *Knowledge-based systems in artificial intelligence*. New York: McGraw-Hill.

DeMulder, R.V. and Gubby, H.M. 1983. Legal decision making by computer: An experiment with sentencing, *Computer/Law Journal* IV, 243–303

Denning Lord Alfred Thompson 1953. Royal Commission on Capital Punishment, 1949–53 Report, Cmd. 8932, London: Her Majesty's Stationery Office.

Dewey, J. 1924. Logical method and law, *Cornell Law Quarterly* 10, 17–27.

Dick, J.P. Representation of legal text for conceptual retrieval. In *Proceedings of the Third International Conference on Artificial Intelligence and Law (ICAIL-91)*, Oxford: ACM Press, pp. 244–253.

Doob, A.N. and Park, N.W. 1987. Computerized sentencing information for judges, *Criminal Law Quarterly* 30, 54–72.

Dreyfuss, H.L. 1992. *What Computers Still Can't Do*. Cambridge, Mass.: MIT Press.

Dyer, M.G. 1983. *In-depth understanding*. Cambridge, Mass.: MIT Press.

Forbus, K.D. 1988. Quantitative physics: Past, present and future. In H. Shrobe (ed.), *Exploring artificial intelligence*. San Mateo, California: Morgan Kaufmann.

Fox, J.A. 1981. *Models in quantitative criminology*. New York: Academic Press.

Freed, S.J. 1992. Federal sentencing in the wake of guidelines: unacceptable limits of the discretion of sentences. *Yale Law Journal* 101(8): 1681–1754.

Gardner, A. 1987. *An artificial intelligence approach to legal reasoning*. Cambridge, Mass.: MIT/Bradford Books.

Garland, D. 1991. Sociological perspectives in punishment. In M. Tonry (ed.), *Crime and justice – a review of research*, Vol. 14. Chicago: University of Chicago Press, pp. 115–165.

Gentner, D. 1989. *Finding the needle: Accessing and reasoning from prior cases. In C. Hammond (ed.), Proceedings: Workshop on case-based reasoning (DARPA)*. San Mateo, California: Morgan Kaufmann, pp. 137–143.

Ginsberg, A. 1993. A unified approach to automatic indexing and information retrieval, *IEEE Expert* 8(5), 46–56.

Greenleaf, G. 1991. Information technology and the law, *Australian Law Journal* 65: 45–48.

Hafner, C.D. 1978. *An information retrieval system based on a conceptual model of legal knowledge*. Ph.D. Thesis. Ann Arbor, Michigan: University of Michigan.

Hafner, C.D. 1987. Conceptual organization of case law knowledge bases. In *First International Conference on Artificial Intelligence and Law (ICAIL-87)*. Boston: ACM Press, pp. 35–42.

Hart, H.A.L. 1963. *Law liberty and morality*. Oxford: Oxford University Press.

Hassett, P. 1993. Can expert system technology contribute to improved bail conditions, *Int. J. of Law and Information Technology* 1, 144–188.

Hassett, P. 1994. Private communication.

Hogarth, J. 1988. *Computer and the law: Sentencing database system, user's guide*. Vancouver: LIST Corporation.

Home Office White Paper 1996. *Protecting the public: The government's strategy on crime in England and Wales*. London: Her Majesty's Stationery Office.

Hudson, B. 1987. *Justice through punishment – A critique of the 'justice' model of correction*. London: Macmillan.

Hutton, N., Patterson, A., Tata, C., and Wilson, J. 1995. Decision support for sentencing in a common law jurisdiction, *Fifth International Conference on Artificial Intelligence and Law (ICAIL-95)*. Washington D.C.: ACM Press, pp. 89–95.

Johnson, L. and Keravnou, E.T. 1985. *Expert systems technology*. London: Abacus Press.

Kass, A.M., Leake, D.B., and Owens, C.C. 1986a. Swale: A program that explains. In R.C. Schank (ed.), *Explanation patterns: Understanding mechanically and creatively*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, pp. 232–254.

Kass, A.M. 1986b. Modifying explanations to understand stories, *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*. Amherst, MA.

Kass, A.M. 1989. Adaption-based explanation: Extending script/frame theory to handle novel input, *Proceedings of IJCAI-89*. San Mateo, California: Morgan Kaufmann, pp. 141–147.

Kass A.M. 1990. Developing creative hypotheses. *Technical Report #6*. Northwestern University, Illinois: Institute for the Learning Sciences.

Kolodner, J.L. 1981. Organization and retrieval in a conceptual memory for events, *Proceedings of IJCAI-81*. San Mateo, California: Morgan Kaufmann, pp. 227–233.

Kolodner, J. 1983. Maintaining organization in a conceptual memory for events, *Cognitive Science* 7: 281–328.

Kolodner, J.L., Simpson, R.L., and Sycara-Cyranski, K. 1985. A computer model of case-based reasoning in problem solving. *Proceedings of IJCAI-85*. San Mateo, California: Morgan Kaufmann, pp. 284–290.

Kolodner, J.L. 1988. Retrieving events from a case memory: A parallel implementation, *Proceedings of a Workshop on CBR*. San Mateo, California: Morgan Kaufmann, pp. 233–249.

Kolodner, J. 1993. *Case-based reasoning*. San Mateo, California: Morgan Kaufmann.

Leake, D. 1992. *Evaluating explanations: A content theory*, Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Lehnert, W., Dyer, MG., Johnson, P, Yang, C., and Harley, S. 1983. BORIS – An experiment in in-depth understanding of narratives, *Artificial Intelligence* 20, 15–62.

Lovegrove, A. 1987. An evaluation of judicial models for sentencing guidelines In D.C. Pennington and S. Lloyd-Bostock (eds.), *The psychology of sentencing: Approaches to consistency and disparity*. Oxford: Oxford University Press, pp. 133–142.

Lovegrove, A. 1989. *Judicial decision making sentencing policy and numerical guidance*. Berlin: Springer Verlag.

Martinson, R. 1974. What work? – Questions and answers about prison reform', *The Public Interest* 35, 22–54.

Mital, V., Stylianou, A., and Johnson, L. 1991. Conceptual information retrieval in litigation support systems. In: *Proceedings of the Third International Conference on Artificial Intelligence and Law (ICAIL-91)*. Oxford: ACM Press, pp. 235–243.

Murbach, R. and Nonn, E. 1993. Similarity in harder cases. *Proceedings of the Fourth International Conference on Artificial Intelligence and Law (ICAJL-93)*. Amsterdam: ACM Press, pp. 236–244.

Posner, R. 1985. An economic theory of criminal law, *Columbia Law Review* 85, 1193–1231.

Redmond, M. 1990. Distributed cases for case-based reasoning; facilitating use of multiple cases, *Proceedings of AAAI-90*. Boston: American Association for Artificial Intelligence, pp. 304–309.

Reynolds, D. and Beck, T. 1993. Tennessee offender management information system, *AI Magazine* 14(3): 61–68.

Riesbeck, C.K. and Schank, R.C. 1989. *Inside case-based reasoning*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Rissland, E.L. and Ashley, K.D. 1987. A case-based system for trade secrets law, *First International Conference on Artificial Intelligence and Law*. Boston: ACM Press, pp. 60–66.

Rissland, E.L. and Friedman, M.T. 1995. Detecting change in legal concepts. In: *Fifth International Conference on Artificial Intelligence and Law (ICAIL-95)*. College Park, MD: ACM Press, pp. 127–136.

Salton, G. 1989. *Automatic Text Processing*. Reading, Mass.: Addison-Wesley.

Schank, R.C. and Abelson, R.P. 1977. *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Schank, R.C. 1980. Language and memory, *Cognitive Science* 4(3): 243–284.

Schank, R.C. 1982. *Dynamic memory*. Cambridge: Cambridge University Press.

Schank, R.C. 1986. *Explanation patterns*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Schank, R.C. and Leake, D.B. 1989. Creativity and learning in a case-based explainer, *Artificial Intelligence* 40: 353–385.

Schank, R.C. 1990. *Tell me a story – a new look at real and artificial memory*. New York: Charles Scribner's Sons, Macmillan Publishing Company.

Schank, R.C., Kass, A., and Riesbeck, C.K. (eds.) 1994. *Inside case-based explanation*, Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Schild, U.J. 1992. *Expert Systems and Case Law*. Chichester: Ellis Horwood.

Schubert, L.K. and Hwang, C.E. 1989. An episodic knowledge representation for narrative texts, *First International Conference on Principles of Knowledge Representation and Reasoning*. Toronto, Canada, pp. 444–458.

Schweighofer, E., Winiwarter, W., and Merkl, D. 1995. Information filtering: The computation of similarities in large corpora of legal texts. In: *Fifth International Conference on Artificial Intelligence and Law (ICAIL-95)*. College Park, MD: ACM Press, pp. 119–126.

Sergot, M.J. 1991. The representation of law in computer programs. In T. Bench-Capon (ed.), *Knowledge-based systems and legal applications*. London: Academic Press, pp. 3–67.

Shapira, M. 1990. Computerized decision technology in social service, *International Journal of Sociology and Social Policy* 10, 138–164.

Simon E. and Gaes, G. 1989. ASSYST – computer support for guideline sentencing, *Second International Conference on Artificial Intelligence and Law (ICAIL-89)*. Vancouver: ACM Press, pp. 195–200.

Skalak, D.B. and Rissland, E.L. 1992. Arguments and Cases, *Artificial Intelligence and Law* 1, 3–44.

Smith, J.C., Gelbart, D., MacCrimmon, K., Atherton, B., McClean, J., Shinehoft, M., and Quintana, L. 1995. Artificial intelligence and legal discourse: The flexlaw legal text management system, *Artificial Intelligence and Law* 3(1–2): 55–95.

Sowa, J.F. 1984. *Conceptual structures: Information processing in mind and machine*. Reading, Mass.: Addison-Wesley.

Susskind, R. 1987. *Expert systems in law*. Oxford: Oxford University Press.

Thomas, D.A. 1979. *Principles of sentencing.* London: Heinemann, 2nd ed.

Thomas, D.A. 1982. *Current sentencing practice*. London: Sweet and Maxwell.

Tonry, M. 1993. *Sentencing Commissions and their Guidelines*. In M. Tonry (ed.), *Crime and Justice – A Review of Research*. Chicago: Chicago University Press, 17: 137–195.

Tonry, M. 1993a. The failure of the U.S. Sentencing Commission's guidelines, *Crime and Delinquency* 131.

Toulmin, S.E. 1958. *The uses of argument*. Cambridge: Cambridge University Press.

Turtle, H. 1995. Text retrieval in the legal world, *Artificial Intelligence and Law* 3(1–2), 5–54.

Tversky, A. 1977. Features of similarity, *Psychological Review* 84(4), 327–352.

van Noortwijk, C., Piepers, P.A.W., van der Wees, J.G.L., and De Mulder, R.V. 1991. The JURICAS system: New applications and future developments. In *Proceedings of the Third International Conference on Artificial Intelligence and Law (ICAIL-91)*. Oxford: ACM Press, pp. 201–206.

von Hirsch, A. 1992. *The Swedish Sentencing Law*. In A. von Hirsch and A. Ashworth (eds.), *Principled Sentencing*, Boston: Northeastern University Press, pp. 292–307.

von Hirsch, A. and Ashworth, A. 1992. *Principled Sentencing*. Boston: Northeastern University Press.

von Hirsch, A. 1993. *Censure and Sanctions*. Oxford: Oxford University Press.

Weatherburn, D., Crettenden, I., Bray, R., and Poletti, P. 1988. *New South Wales sentencing information system*. New South Wales: Judicial Commission of New South Wales.

Wellbank, M. 1983. *A review of knowledge acquisition techniques for expert systems*. Ipswich: British Telecom.

Wilkins, L.T., Kress, J.M., Gottfredson, D.M., Calpin, J.C., and Gelman, A.M. 1978. *Sentencing guidelines: Structuring judicial discretion*. Washington D.C.: United States Department of Justice.

Wilson, J. 1996. Personal communication.

Zeleznikow, J. and Hunter, D. 1994. *Building Intelligent Legal Information Systems*. Deventer, the Netherlands: Computer/Law Series, KIuwer Law and Taxation Publishers.