THE GEOMETRY OF CASE LAW

Using seven well known cases dealing with the right to possession of movable property by a finder, Alan Tyree explores some aspects of the use of mathematical techniques in the realm of case law.

I. JURIMETRICS: AN OVERVIEW

The term "jurimetrics" was introduced by Lee Leovinger in 1948.[1] As a branch of jurisprudence, its origins have been traced to the movement known as American Realism.[2] Indeed, both have claimed to pursue a scientific study of jurisprudence.

However, jurimetrics, as the name implies, is really concerned with the application of quantitative and mathematical methods to legal problems. The development of such a discipline was, I suggest, a product of the intellectual climate of the post-war years. In the years before the war the discovery was made that mathematics, often thought to be of use only to the "precise" sciences, could be of service to the social sciences as well.[3] The needs of the war gave impetus to this movement with the result that quantitative and mathematical methods came to be seen as natural and powerful tools for the advancement of such disciplines as psychology and sociology. In such an intellectual climate, the development of jurimetrics must have been inevitable.

The results of jurimetrics have not been spectacular to date. The main streams of development will be reviewed below, but the impact of jurimetrics must surely be a disappointment to the early writers. There are, I suggest, several reasons for this slow development.

At a practical level, there has never been a large source of manpower trained in both legal and mathematical studies. The system of secondary and tertiary education of the type found in New Zealand virtually guarantees that students interested in the study of law will be unfamiliar with the principles of modern mathematics, and vice versa. There is supporting evidence for this statement in the observation that jurimetrics has been primarily an American phenomenon.[4] The practice in that country of a broad four year university degree followed by a training in law has meant that there are at least some lawyers and, more importantly, some law teachers with at least an undergraduate training in mathematics.

---

1. Loevinger, "Jurimetrics: The Next Step Forward" (1948) 33 Minn. L. Rev. 755.
2. Lord Lloyd of Hampstead, *Introduction to Jurisprudence* (3rd ed. 1972) 415.
3. Newman, *The World of Mathematics* (New York, 1956).
4. There are two American journals devoted to the subject: Jurimetrics Journal (formerly Modern Uses of Logic in Law) and Rutgers Journal of Computers and Law.

The social sciences, concerned as much with the training of researchers as with the training of practitioners, have avoided this problem. Quantitative and mathematical methods have been incorporated into the curriculum of these studies. The result has been self reinforcing. As more social scientists became familiar with the methods, they used these methods in their own research. This in its turn led to a greater demand for such methods to be incorporated into the curriculum. Such a development might be much slower in legal education because of its professional orientation.

The excitement of the new discipline led also, I suggest, to some ideological errors. Loevinger's original paper is typical of many in seeming to envisage the replacement of traditional jurisprudence with jurimetrics.[5] The analogous replacement in other disciplines has not occurred, nor has it even been thought to be desirable. This ideological excess has had practical repercussions. The claim that mathematics can solve all jurisprudential problems has met the predictable response that it can solve none of them.[6]

But the accomplishments of jurimetrics, if not spectacular, have not been negligible. I shall discuss three main areas: the use of automatic information retrieval, the analysis of court decisions by the behaviouralists, and the fact content analysis of court decisions.[7]

A. Automatic Information Retrieval

It has become commonplace to remark upon the explosion of legal literature, and as in many other areas, the growth seems to be exponential with little reason to suppose that it will slow in the foreseeable future. In addition, law, particularly case law, suffers from a problem which does not trouble most scientific disciplines, namely the long useful life of documents.[8]

In this situation, it is not surprising that computerised library systems should be suggested. There are already a few operational systems in the world and the number seems bound to increase.[9]

There has been surprisingly little research on the question of necessity of the new methods. Critics may argue that such systems are unnecessary luxuries or even toys which are positively detrimental to good library and research habits. They may further argue that law has a system of indexing which is superior to any other discipline and that

---

5. Loevinger, op. cit.
6. Weiner, "Decision Prediction by Computer: Nonsense Cubed-and Worse" (1962) 48 A.B.A.J. 1023, and answering comments by Kort, "Simultaneous Equations and Boolean Algebra in the Analysis of Judicial Decisions" (1963) 28 Law and Contemporary Prob. 143.
7. These are not the only areas of interest, but they are the most highly developed.
8. Tapper, *Computers and the Law* (London, 1973), 112; Diamond, "Codification of the Law of Contract" (1968) 31 M.L.R. 361.
9. Tapper, op. cit., ch. 7.

the so-called flood of literature has not stopped either practising or academic lawyers from functioning.

But the research that has been done suggests that all is not as well as might be thought.[10] The effectiveness of any document retrieval system, be it computer or the standard manual use of indexes, is usually measured by two indices. In any library search, the documents may be divided into groups, those which are relevant to the search and those which are not.[11] An ideal search would retrieve only those which are relevant. It would retrieve all of these and no others. *Recall* is the percentage of all of those relevant documents which are actually retrieved by the system. *Precision* is the percentage of relevant documents among those actually retrieved. High recall is thoroughness. High precision avoids the handling of irrelevant material.

The most promising experiments in the retrieval of case law by computers use the so-called full text scanning. The entire text of a reported case is stored in some machine readable form. Various retrieval algorithms are used, but all rely upon the notion that meaning is to be found by the occurrence of certain words or word combinations. The most comprehensive work of this sort has been done by Salton on a collection of engineering documents.[12] Requests to the system are in the form of a question in natural language.

Colin Tapper has used these standard methods in some experiments with a collection of case law.[13] He compared the automatic retrieval system with the performance of trained lawyers using the standard library research methods. His results indicated that the computer system offered high recall (70%) but low precision (29%). The manual searches for the same problem produced relatively low recall (49%) with astonishingly high precision (92%).

These figures for the automatic systems will certainly be increased considerably, but even they indicate that a computer system would have value for academic research, where recall may be more important than precision.

B. The Behaviouralists

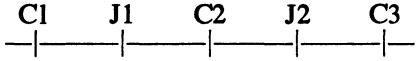
The leading writer in this area has been Glendon Schubert. His writings are voluminous and include four books which are required reading for those interested in the area: *Judicial Behaviour* (Chicago, 1964), *Judicial Decision Making* (New York, 1963), *The Judicial Mind* (Evanston, 1965), and *The Judicial Mind Revisited* (New York, 1974).

The behaviouralists focus on "attitudinal" variables. The early studies were one dimensional and these provide the easiest introduction

10. Tapper, op. cit.
11. Salton, *Automatic Information Organisation and Retrieval* (New York, 1968).
12. Salton, op. cit.
13. Tapper, op. cit.

to the subject. The basic model supposes a scale of values, e.g. a civil liberties scale. The assumption is that each judge has formed attitudes toward civil liberties which enable him to be represented by a point on this scale. Cases are then thought of as stimuli to which the judge responds by a vote, i.e. he decides the case. The notion is that cases lying to the left of the point representing the judge will be decided one way by him, those to the right in the other way.

For example, suppose that judges J1 and J2 and civil liberties cases C1, C2 and C3 may be represented by the following configuration:



According to the model, both judges would agree on cases C1 and C3, one being decided in a manner extending civil liberties and the other in a manner restricting civil liberties. The case C2 would be a split decision.

This basic model has been replaced by a more sophisticated multi-dimensional one; judges and cases are positioned not on a line but in a multi-dimensional "attitudinal" space.

The behaviouralists work has centred upon relatively large policy-oriented courts, the Supreme Court of the United States being the usual object of study. The main reason for this, apart from the natural interest of American writers in such courts, is that the method used to position the judges in attitudinal space requires both a large number of judges and a large number of split decisions.[14] Courts satisfying these methodological requirements are invariably policy making constitutional courts.

But such courts are not the only ones for which the results of an attitudinal study would be interesting. I suggest that there are many studies which are interesting precisely because they seem to indicate differing attitudes among judges. Thus, for example, Frank mentions a study concerning sentencing in cases of intoxication and vagrancy which showed wide variation in the practice of judges in Chicago.[15] I suggest that the reason that such variation alarms us is that we intuitively believe that the facts do not vary much from case to case in such an area. Consequently, the variation must be caused by the attitudes of the Judges towards intoxication and vagrancy. Here the attitudes of judges are of interest because they tell us something about our system of administering justice. This is quite different from the reason for our interest in the attitudes of judges in a policy making constitutional court, but no less vital.

---

14. The behaviouralists originally relied upon Guttman Cumulative Scaling. This has been largely superseded by multi-dimensional scaling using correlations between judges as input. See section *C*.
15. Frank, *Law and the Modern Mind* (Gloucester, Mass., 1970). The study referred to is Everson, "The Human Element in Justice" 10 J. of Crim. Law and Criminology 90.

A method of conducting a behaviouralist study which does not depend upon analysing split decisions of a large court is suggested in part II of this paper.

Schubert's work has, as might be expected, provoked a lot of comment.[16] Schubert is quite capable and extraordinarily willing to answer most of the criticism himself. Briefly, the criticisms fall under two headings, criticism of the mathematical techniques and criticism of the basic model.[17]

The faults of the mathematical techniques used in the early studies largely disappear with the use of the more sophisticated methods, particularly the use of non-metric multi-dimensional scaling. Since these newer methods may be used to provide essentially the same results, the force of the criticism would seem to be now spent.

Criticism of the basic model can only be answered by demonstrating the usefulness of the model itelf. Testing this usefulness involves, among other things, testing hypotheses concerning the model and using the model to make predictions which may then be tested against reality. Since this is precisely what Schubert's writings are all about, the interested reader must be referred to them.

C. *Fact Content Analysis*

These studies attempt to use the factual content of cases as the raw material for study. Consequently, they are open to the criticisms of the fact sceptics who assert that the facts of the case are essentially a construction of the judge. The fact content analysts attempt to meet this criticism in two ways. Firstly, they say that even if it is true that the "facts" that appear in the judgments are artificial, they are nevertheless important for analysing a series of cases and for predicting what the future constructs, and consequently the future decisions, will be. Secondly, they point to the consistency observed in their studies and the relative success of their predictions based on fact content to argue that the fact sceptics have overstated the case.[18] While there may often be some uncertainty, their results suggest that the traditional importance of facts is not to be ignored.

Once the factual content of a series of cases has been determined, it may be used in several different forms of analysis. Kort has used factor analysis followed by regression in an attempt to predict in a statistically optimal fashion the outcome of cases.[19] Lawlor has used

16. Schubert, *The Judicial Mind Revisited* (New York, 1974).
17. For example, Tanenhause, "Cumulative Scaling of Judicial Decisions" (1966) 79 Harv. L.R. 1583.
18. Lawlor, "Axioms of Fact, Polarization and Fact Ranking — Their Role in Stare Decisis (1969) 14 Vill. L.R. 703. Also, Lawlor "Personal Stare Decisis" (1968) 41 S. Cal. L.R. 73.
19. Kort, "Predicting Supreme Court Decisions Mathematically: A Quantitative Analysis of the 'Right to Counsel Cases' (1957) 51 Am. Pol. Sci. Rev. 1; and "Simultaneous Equations and Boolean Algebra in the Analysis of Judicial Decisions (1963) 28 Law and Contemporary Prob. 143.

methods closely related to Boolean algebra and symbolic logic to study the consistency of a line of cases.[20]

The fact content analysts have emphasised the prediction of future court decisions. They have been criticised for this, the criticism focusing on what is imagined to be a conflict between "understanding" and "prediction". A very good example of this type of criticism is Professor Fuller's discussion.[21] He suggests that both behaviouralists and fact content analysts would profit by concentrating on "understanding" which he believes is more important than "prediction". The example that he uses to strengthen his argument is enlightening. The Darwinian theory of evolution, says Fuller, is an excellent example of a theory that greatly increased our understanding without offering or pretending to predict. Professor Fuller failed to observe, however, that as the Darwinian theory has developed to a point where it is capable of making predictions, i.e. testable hypotheses, it has been found that we understand much less than we imagined, that indeed the Darwinian theory is insufficient to account for the observed changes in life forms.[22] It would seem to be a general rule that knowledge progresses by making guesses which must then be formulated as theories which must then be tested against their predicted consequences. There is no reason to suppose that understanding and prediction are in any way opposed.

Furthermore, it is the reliable prediction of court decisions which would be of greatest value to the law practioner. Prediction with 100% accuracy is out of the question, but predictions of the form "X will win this case with a probability Y%" are probably within the ultimate scope of current methods.[23]. Such information, if reliable, could be of value to practitioners when advising clients.

II. Similarity Measures on Cases

A. Similarity in English Law

Cross opens his book *Precedent in English Law* with the following sentence: "It is a basic principle of the administration of justice that like cases should be decided alike." In English law, this principle takes the form of the doctrine of precedent, which in turn holds that it is the ratio decidendi of a case which is binding.

The difficulty in abstractly defining the ratio decidendi is well known, but there is agreement that the factual content of the case is

---

20. Lawlor, op. cit.
21. Fuller, "An Afterword: Science and the Judicial Process" (1966) 79 Harv. L.R. 1604.
22. Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. V: Darwinian, Neo-Darwinian, and Non-Darwinian Evolution (Berkeley, 1972).
23. Cowan, "Decision Theory in Law, Science and Technology" (1963) 17 Rutgers L. Rev. 499.

important. Cross cautions that "Judgments must be read in the light of the facts of the cases in which they are delivered".[24] The important position of facts has been used by Dr. Goodhart to give a lucid definition of the ratio decidendi of a case.[25]

This emphasis upon the factual content of a case finds its way into modern legal education. Legal students are taught the importance of carefully analysing the factual situation. Given a problem to resolve, the student searches for cases which have "similar" fact situations. If he is to argue for a particular outcome, he will use the similar case either to bolster his argument or he will seek to show that although the case appears to be similar there are features which make it quite different from his factual situation.

This is not necessarily to adopt a jurisprudential point of view which holds that cases are decided on the basis of the facts and the facts alone. The point that the "facts" may not appear fully until discovery of the court's opinion is well taken,[26] but in the context of court practice the point is irrelevant. Regardless of the jurisprudential theory, lawyers must be able to deal with cases in the manner described in the preceding paragraph. Whether the process is thought of as the "real" course of the law or merely some sort of procedural gloss which covers the "true" underlying process makes no practical difference.

All of this suggests that there is a notion of similarity between pairs of cases which is based upon the factual content of the cases. If this notion is to be useful, then it must be shown to be relatively stable; lawyers should have similar notions of similarity between cases. This is investigated in section *B.* of this part, where it is shown that in a particular case that there is a high degree of concordance. This high degree of concordance in turn suggests that it might be possible to quantify the notion of case-case similarity; this quantification is investigated in section *C.* Two uses for the quantification are the subject of section *D.*

B. Is there a Common Notion of Similarity?

The above quotation from Cross seems to presuppose that there is a relatively common notion of similarity between cases. This presupposition does not appear to be unreasonable given the relatively uniform cultural and educational background of lawyers.

Concordance among four subjects ranking twenty eight pairs of cases was tested. The subjects were students who had passed LAWS 101: The Legal System at Victoria University of Wellington during the 1975 session. Material used in that course included a series of so-called "finders" cases. Students in the course were expected to be

24. Cross, *Precedent in English Law* (1961), p. 39.
25. Goodhart, *Essays in Jurisprudence and the Common Law,* The definition is lucid, but it is not beyond criticism: Stone, "The *Ratio* of the *Ratio Decidendi*" (1959) 22 M.L.R. 597.
26. Frank, *Law and The Modern Mind* (Gloucester Mass., 1970).

familiar with the cases and to have formulated ideas about the law represented by the cases. They were required to apply this knowledge to a lengthy factual type of question on the final examination.

The cases used for the experiment were: *Armory* v. *Delamirie*,[27] *Bridges* v. *Hawkesworth*,[28] *Elwes* v. *Brigg Gas Co.*,[29] *South Staffordshire Water Co.* v. *Sharman*,[30] *Hannah* v. *Peel*,[31] *Corporation of London* v. *Appleyard*,[32] and *Moffatt* v. *Kazana*.[33] The case of *Corporation of London* v. *Appleyard*, however, is treated as two separate cases: *Corporation of London* v. *Yorkwin*, and *Yorkwin* v. *Appleyard*. This is a total of eight cases, giving twenty eight pairs.

Each pair of cases was written on a separate card and the cards were shuffled. The subject was given the deck of cards with instructions to order the deck according to their own notions of which pairs were most similar. The resulting rankings are shown in Figure 1.

For reasons of clarity, the cases are represented by letters in all of the Figures 1-6 as follows:

A: *Armory* v. *Delamirie*
B: *Bridges* v. *Hawkesworth*
C: *Elwes* v. *Brigg Gas Co.*
D: *Hannah* v. *Peel*
E: *Corporation of London* v. *Yorkwin*
F: *Moffatt* v. *Kazana*
G: *South Staffordshire Water Co.* v. *Sharman*
H: *Yorkwin* v. *Appleyard*

Looking at the table in Figure 1, it can be seen that the four subjects did not agree perfectly in their ranking of the twenty-eight case pairs. Nevertheless, there is by no means total disagreement. For example, all four subjects rated the pair *Bridges* v. *Hawkesworth* – *Hannah* v. *Peel* very highly similar, while the pair *Armory* v. *Delamirie* – *Moffatt* v. *Kazana* is ranked near the end of the list by all subjects.

There is a numerical measure of the degree of agreement among rankings. Devised by the statistician M. G. Kendall, it is called Kendall's coefficient of concordance and is usually denoted by W.[34] If all of the rankings are in perfect agreement, then $W=1$. If there is no agreement at all, then $W=0$. In cases of imperfect agreement, W will be a number between 0 and 1 which will measure the degree of agreement among the subjects.

When W is calculated for the rankings of the case pairs given by the four subjects, it is found to have a value of .83.

27. (1721) 1 Strange 505.
28. (1851) 21 L.J.Q.B. 75.
29. (1886) 33 Ch. D. 562.
30. [1896] 2 Q.B. 44.
31. [1945] 1 K.B. 509.
32. [1963] 1 W.L.R. 982.
33. [1969] 2 Q.B. 152.
34. Kendall, M.G., *Rank Correlation Methods* (2nd ed. London, 1955).

Figure 1: Rankings of the 28 pairs of cases by the four subjects and the ranking generated by the metric.

Case pairs	Subjects 1	2	3	4	5
A – B	24	16	21	23	9
A – C	26	22	28	25	27
A – D	21	15	23	24	15
A – E	25	24	15	26	23
A – F	28	28	27	28	28
A – G	27	20	26	27	24
A – H	23	23	25	19	19
B – C	5	11	9	11	11
B – D	2	1	3	3	2½
B – E	3	21	11	14	16
B – F	17	26	20	22	22
B – G	7	8	6	9	7
B – H	20	13	16	21	11
C – D	8	9	5	6	5½
C – E	4	6	2	1	2½
C – F	11	17	18	20	20½
C – G	1	2	8	5	8
C – H	10	3	12	10	13
D – E	14	12	7	8	11
D – F	16	27	14	15	18
D – G	9	7	4	4	4
D – H	19	10	10	12	5½
E – F	12	25	24	16	26
E – G	13	5	13	7	14
E – H	22	14	17	13	17
F – G	15	19	22	18	25
F – H	18	18	19	17	20½
G – H	6	4	1	2	1

Is .83 a high degree of agreement? Statisticians answer the question in the following way: let us suppose that the four subjects were really just lazy oafs with no interest in the experiment at all. In order to save time, they just reshuffle the cards and return them to the experimenter. It is certainly possible, though unlikely, that the results could come back exactly as they appear in Figure 1, or in some other arrangement which would show an agreement which of at least $W=.83$. Just how likely is such an event? Calculations show that it has a probability of less than 1%. Such an unlikely possibility leads us to conclude that $W=.83$ indicates a high degree of agreement among the subjects.

Statisticians summarize the situation by saying that $W=.83$ is statistically significant at the 1% level.

There is a second way in which agreement may be numerically

measured. The subjects may themselves be compared pair-wise by means of the Spearman rank correlation coefficient.[35] This coefficient ranges from +1 when there is perfect agreement between a pair of subjects to a minimum of —1 in the case where one ordering is the reverse of the other. The Spearman coefficients are shown in Figure 2. All are statistically significant at the 1% level.

Figure 2: Spearman rank correlation coefficients.

Subject	1	2	3	4	D
1	1	.65	.80	.83	.63
2		1	.77	.80	.84
3			1	.90	.88
4				1	.80
D					1

These results seem to vindicate the supposition of an underlying notion of similarity between cases. There are, however, two effects that require further testing.

Firstly, the coefficients may have been depressed by forcing the subjects to make a complete ranking of all twenty-eight pairs of cases. All of them found the job difficult and reported that many of the choices, particularly "in the middle", were somewhat arbitrary. If this is correct, it would have a randomizing effect on the rankings which would depress the coefficients. It would be interesting to repeat the experiment using the dichotomous (similar/not similar) or perhaps a trichotomous (similar/somewhat similar/not similar) classification.

Secondly, the coefficients would tend to be inflated by the common background of the subjects. All of the subjects met the cases for the first time in the LAWS 101 course; they studied the cases under the guidance of the same teacher.

It would require further testing to know the magnitude of these effects.[36]

C. *Quantification*

As was seen in section *B*, there is some evidence of the existence of a relatively common notion of similarity between cases. The purpose of this section is to capture this notion in a quantitative form. The usefulness of such a quantification will be discussed in section *D*.

For technical reasons, it is more convenient to think in terms of dissimilarity. For a given pair of cases C1 and C2, a number will be

35. Kendall, M.G., op. cit.
36. I expect these effects to be small individually. The net effect will, of course, be smaller.

assigned which is a measure of their dissimilarity. If the two cases are factually identical, then this number will be zero, reflecting the situation that lawyers refer to as "being on all fours". In all other cases, the number should be positive indicating a degree of dissimilarity. The measure of dissimilarity proposed here has all of the properties which make it what mathematicians call a metric.[37] In accordane with mathematical terminology, the dissimilarity number assigned to a pair of cases will be called the distance between them.

Rather than discussing the construction of the metric in an abstract form, the finders cases of section *B.* will be used as illustration.

The construction process is a three stage one. Firstly, a list of questions intended to reflect the fact content of the cases is proposed.[38] The essential features of these questions are that they should elicit all relevant facts and that their answers should be as objective as possible. There is an unavoidable element of subjectivity in the choice of the questions.[39] It is hoped, however, that the questions are precise enough to be answered objectively. As will be seen, the objectivity in the answers is an important consideration in the construction of the metric.

Secondly, the answers to the questions are coded. A "yes" answer is coded as a "1" and a "no" answer is coded as an "0". Each case may then be considered to be represented by a test profile, a string of 0's and 1's which represent the answers to the questions. If the questions are adequate, the test profile should represent faithfully the factual content of the case.

Thirdly, the metric is constructed using the test profiles of the cases.

There is one further consideration. Some facts seem to be more important in context than others. For example, in the finders cases, the case of *Moffatt* v. *Kazana* seems to stand out because the true owner of the lost chattel is one of the litigants. In this particular collection of cases, that fact seems of great importance when comparing *Moffatt* with any of the other cases. The metric will be constructed so as to reflect this inequality of importance.

The questions used are:

Q1: Did the finder control the real estate where the chattel was found?

Q2: Was the chattel attached to the real estate?

---

37. Kelly, J.L., *General Topology* (New York, 1955). In general, the defined function will only be a pseudo-metric since distinct cases may have zero distance. This does not in fact occur with the eight cases used here.
38. Compare these with the factors proposed by Harris, "The Concept of Possession in English Law" in *Oxford Essays in Jurisprudence* (1961), Ed. Guest.
39. This element of subjectivity should not be over emphasised. There are indications that this procedure is less subjective than the ordinary system of indexing: Lawlor, "Fact Content Analysis of Judicial Opinions" (1968) 8 Jurimetrics Journal, 107.

Q3: Did the other claimant have title to the real estate?
Q4: Did the other claimant have title to the chattel?
Q5: Was there a prior legal agreement concerning the chattel?
Q6: Did either party rely upon a lease?
Q7: Is there a master/servant relationship between the other claimant and the finder?
Q8: Was the chattel in a position so as to be difficult to find?
Q9: Was there an attempt to find the true owner of the chattel?
Q10: Had either party knowledge of the chattel prior to the finding?

The coded answers to the questions are displayed in Figure 3.[40]

Figure 3: Answers to Questions.

	A	B	C	D	E	F	G	H	AVE	VAR
Q1	0	0	1	0	0	0	1	1	$\frac{3}{8}$	.234
Q2	0	0	1	1	0	1	0	1	$\frac{1}{2}$	.25
Q3	0	1	1	1	1	0	0	1	$\frac{5}{8}$	.234
Q4	0	0	0	0	0	0	1	0	$\frac{1}{8}$	.109
Q5	1	0	0	0	0	0	0	1	$\frac{1}{4}$	.188
Q6	0	0	1	0	0	0	0	1	$\frac{1}{4}$	.188
Q7	0	0	0	1	0	1	0	0	$\frac{1}{4}$	.188
Q8	0	0	1	1	1	1	1	1	$\frac{3}{4}$	.188
Q9	0	1	1	1	1	1	1	1	$\frac{7}{8}$	.109
Q10	0	0	0	0	0	0	1	0	$\frac{1}{8}$	.109

The list of questions is clearly not exhaustive of the questions that might have been asked, nor are the answers as objective as might be desired. However, for this pilot study, they will be seen to be adequate. More questions with less ambiguity should improve the performance of the resulting metric.

Figure 3 shows two additional numbers associated with each question; the average, denoted by AVE, and the variance, denoted by VAR.

The average for a question is just the arithmetic average of its coded responses.

The variance of a question is a measure of lack of uniformity of response. If the responses to a question are all the same, its variance is zero. If the answers are evenly divided, then the variance is at a maximum. In the special circumstances of Figure 3, the variance may

40. If it was impossible to answer the question from the report, an "0" was coded. An alternative approach, possibly better, would be to select coding at random. Even better would be to use a more complicated coding. There is no mathematical difficulty in this, although the metric is considerably more complicated to describe to non-mathematicians.

be calculated by squaring the average and subtracting that result from the average itself; symbolically, VAR = AVE — $(AVE)^2$.

The distance between any two cases may now be constructed as follows.[41] For each question, compare the responses of the two cases. If the responses are the same, score zero. If the responses are different, score 1/(VAR of the question). When this has been done for each question, add the resulting scores together. This sum is the distance between the cases. Distances between all pairs of cases are shown in Figure 4. For example, to calculate the distance between *Moffatt* v. *Kazana* and *Hannah* v. *Peel*, the only contributions to the score come from questions 1, 3, 4 and 10, since these two cases have the same response to each of the other questions. Thus, the distance D is 1/.234 + 1/.234 + 1/.109 + 1/.109 = 26.90. Note that this figure differs slightly from that given in Figure 4. The latter was calculated to a higher degree of accuracy. The example illustrates the higher weight attached to those questions with a low variance.

Figure 4: Table of distances between cases.

	A	B	C	D	E	F	G	H
A	0	18.74	37.68	24.08	32.34	42.36	33.71	29.14
B		0	18.93	5.33	24.27	32.15	14.67	18.93
C			0	13.60	5.33	31.89	14.93	19.20
D				0	18.93	26.82	9.33	13.60
E					0	37.22	20.27	24.53
F						0	36.15	31.89
G							0	4.27
H								0

The purpose of defining the metric is to capture quantitatively the intuitive idea of similarity/dissimilarity between cases. In order to evaluate the success of this attempt, the metric may be used to rank the twenty-eight case pairs. The cases which are "closest" according to the metric are ranked first and so on. The resulting ranking is also shown in Figure 1.

If the metric has successfully captured the idea of similarity then the ranking generated by the metric should agree well with the rankings of the four subjects. The Spearman rank correlation coefficient of the metric ranking with that of each of the four subjects is shown in Figure 2. These coefficients are all statistically significant at the 1% level.

This substantial agreement with the rankings produced by the four subjects indicates that the metric reflects the intuitive notion of similarity between cases.

41. This is an adaptation of a metric suggested by Kendall and Stuart, *The Advanced Theory of Statistics* (London, 1958) Vol. 3, p. 338.

In addition to the agreement with the human subjects, the metric exhibits a remarkable characteristic which might be called the nearest neighbour rule. Each case has been decided in a manner compatible with its nearest neighbour, i.e. the finder wins case A if, and only if, the finder wins in the case nearest to A. It is to be emphasised that the questions make no reference whatsoever to the outcome of the cases.

D. Uses for the Metric

This section suggests two uses for the measure defined in the preceding section.

1. The geometric representation of cases

In the preceding section, each of the finders cases was represented by its test profile, the coded answers to ten questions relevant to the finders cases. A mathematician would say that each finders case has been represented by a point in ten-dimensional geometric space. Of course, ten-dimensional geometric space is a very unfamiliar concept and cannot be "seen" in our imagination. The question arises if it might not be possible to "reduce" the dimension of the representation to one which we can visualize and, if so, what would be the meaning of such a representation.

An exploratory technique known as non-metric multi-dimensional scaling has been found to be of considerable value in various disciplines.[42] The purpose of the technique is to represent objects (finders cases in the application) as points in a geometric space. Consideration of the geometric properties, e.g. dimension, clustering, etc., then suggests inferences concerning the real properties of the objects represented. It is important to understand that at this stage of development the technique is exploratory and suggestive only. It does not purport to prove the existence of relationships. In spite of these comments, multi-dimensional scaling seems to lead often enough to real relationships so as to inspire some confidence.

The fundamental ideas of multi-dimensional scaling are most easily understood by means of the following example. Suppose that a road map has become separated from the table of mileages between cities. Further suppose that the map has somehow been lost but that the distance chart has been retained. Is it possible to accurately reconstruct the road map from the distance table? The answer is a qualified "yes": if the roads are reasonably straight, a good map may be reconstructed save that the orientation may be lost and the scale of the reconstructed map is arbitrary.[43]

---

42. Shepard, Romney, and Nerlove (Eds), *Multi-dimensional Scaling; Theory and Application in the Behavioural Sciences* (New York, 1972); Green and Carmone, *Multi-dimensional Scaling and Related Techniques in Marketing Analysis* (Boston, 1970).
43. Kendall, D.G., "Maps from Marriages: an Application of non-metric Multi-dimensional Scaling to Parish Register Data" in Hodson, Kendall and Tauto (Eds), *Anglo-Romanian Conference on Mathematics in the Archeological and Historical Sciences, Mamaia 1970* (Edinburgh 1971).

Now consider a slightly more complicated problem. Suppose that it is not known whether the distance table relates to a two dimensional configuration, e.g. a road map, or to a three dimensional system, e.g. a map of a galactic system. Is it possible to determine the dimensionality of the table of distances? Again the answer is "yes". For any such table, it is possible to calculate a number known as the stress for each dimension. The stress is a measure of how well the distance table may be represented by a configuration of points in that dimension. If the representation is exact, the stress will be zero. If the stress is non-zero but still small, the geometric representation is thought of as reasonably reflecting the objects that are to be studied.[44]

Even though there is no a priori reason to suppose that a distance table such as Figure 4 corresponds to a "map" of any dimension, experience has shown that making such a supposition often reveals an underlying structure which is meaningful. The suggestion here is a simple paradigm: if the data will "fit" in, say, two dimensions, then explanation of that data will require precisely two factors or concepts. Other geometric properties of the representation will correspond to other real properties of the data. I know of no logical argument to support this model, but its usefulness cannot be questioned if it is used with caution. This lack of a logical basis should cause no undue alarm, for it is experience, not logic, which is the life of applied mathematics.

The best one dimensional arrangement of the cases is shown in Figure 5.[45] The stress is 20.1%. The best two-dimensional configuration is shown in Figure 6, having a stress of 8.3%.

The stress for the one dimensional configuration reveals that the fit is very poor. This suggests strongly that a single concept will be inadequate to rationalise this series of eight cases. An attempt to do so will probably result in a concept which either does not adequately explain the cases or which is difficult or impossible to apply to actual factual situations.[46]

The one dimensional plot does, however, exhibit a remarkable property: the cases are properly ordered in the sense that those lying to the right of the point marked Θ were all won by the finder and those to the left were all lost by the finder.

This brief analysis of the one dimensional plot may provide some clue as to the popularity of these cases among academic writers.[47] The cases are not properly one dimensional because of the high stress, but they nevertheless exhibit certain one dimensional characteristics, such as the proper ordering. These properties are the geometric reflection of

45. The program used was the Bell Laboratories KYST, based on the algorithm of Kruskal, "Non-metric Multi-dimensional Scaling: A Numerical Method" (1964) 29 Psychometrika 115.
46. Shepard, op. cit.
47. Of these cases, Lord Goddard has said, "These cases, . . . , have long been the delight of professors and text writers, whose task it often is to attempt to reconsile the irreconciliable": *Hibbert* v. *McKiernan* [1948] 2 K.B. 142, 149.

the fact that the cases are difficult to rationalize yet still seem to "hang together", a combination which might be supposed to be very attractive to academic inquirers.

Figure 5: One dimensional plot; co-ordinates from Bell Laboratories KYST program. Stress = 20.1%.

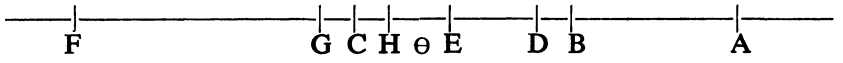
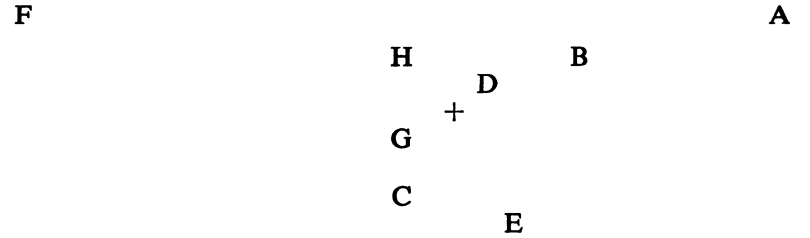


Figure 6: 2-dimensional plot — co-ordinates from Bell Laboratories KYST Program. Stress = 8.3%.



Turning now to the two dimensional plot, it is seen that the fit is good with a stress of 8.3%. This suggests that the eight cases are best explained by the use of two concepts. The search for those concepts and the application of them to further finders cases to test them for validity would be one direction that further research might take.

However, the two dimensional plot reveals very clearly that two of the cases, *Armory* v. *Delamirie* and *Moffatt* v. *Kazana*, are "outliers". This is, of course, in keeping with the observations of other writers, particularly as regards *Armory* v. *Delamirie*. Rather than attempting to explain all of the cases with a two dimensional formulation, an alternate approach is to declare that these two cases are not really finders cases at all. These cases woud then be discarded, the remaining cases rescaled, and the analytic process repeated.

This second direction is more in keeping with current thinking on the finders cases.[48] It will not be pursued here for a technical reason: eight cases already represent a dangerously small number for reliable scaling in two dimensions.[49] The results obtained from the program using only six cases could possibly be quite meaningless.

The low two-dimensional stress also suggests that the banishment of *Armory* v. *Delamirie* from the class of finders cases may have been premature. If all of the cases may be reconciled by the use of two concepts, as suggested by the low two-dimensional stress, then the

48. Tay, "Possession and the Modern Law of Finding" (1964) 4 Sydney L.R. 383; Cohen, "The Finders Cases Revisited" (1970) 48 Tex. L.R. 1001; Riesman, "Possession and the Law of Finders", (1939) 52 Harv. L.R. 1105.
49. Bell Laboratories documentation for KYST.

resulting theory would be of a wider application than a "one-dimensional" theory which would result from the consideration of only a subset of the cases.

Even with this small number of well-known cases, multi-dimensional scaling has shed light on their complexity and structure. At the risk of being repetitive, it must once again be emphasised that multi-dimensional scaling is only, at this stage of development, an exploratory tool. It is a tool which may be used in the study of case law, not as a replacement of the traditional methods, but as a guide to their more efficient use.

2. *A behaviouralist study in New Zealand*

A researcher in New Zealand wishing to carry out a behaviouralist study would face an immediate and frustrating technical obstacle. The only methods available for positioning judges in attidudinal space have required that the judges all be members of the same court and that they produce a reasonably large number of split decisions.

The obstacle is frustrating because it seems so irrelevant to the problem. The behaviouralist hypothesis is as meaningful in New Zealand as it is anywhere, but a methodological problem seems to prevent the testing of that hypothesis.

Fortunately, the obstacle may be overcome. A reading of Schubert's papers will disclose that his use of split decisions is directed toward the production of a measure of similarity between judges. It is this measure of similarity which he uses as input to a multi-dimensional scaling program. The frustrating technical obstacle may be circumvented by finding another method of describing similarity, or dissimilarity, between judges.

A method of doing this will be described using the metric defined in part II *C*. The class of case under consideration must be such that the decisions may be consistently coded. Thus, for example, if appeals against sentence are to be considered, these may be coded by a "1" for allowing the appeal and an "0" otherwise. In this way it makes sense to say that Judge A came to a decision on case B different from that of Judge C on case D.

If Judge A and Judge C disagree on a pair of cases B and D, in the sense described in the preceding paragraph, then calculate the distance between the two cases. Suppose that the distance is d. Score 1/d for the preliminary distance between the judges. The distance between Judge A and Judge C is then defined to be the average value of all such preliminary distances. Notice that if the factual test profiles of the cases are quite different, then the distance between the cases is large and consequently the pair adds little to the overall distance between the judges. If, on the other hand, the distance between the cases is small then there is a more substantial contribution to the distance between the judges.

The table of distances between judges is the substitute for Schubert's similarity measure. These distances may be used as input to a multi-dimensional scaling program and the study then pursued in a manner similar to the studies of Schubert.

III. Conclusions

The metric discussed in the preceding part was shown to capture quantitatively the intuitive idea of similarity between cases. It was indicated how this quantification might be exploited in a New Zealand behaviouralist study.

More interesting is the use of non-metric multi-dimensional scaling to provide a bridge from the quantitative domain to the domain of geometry and intuition. The quantitative information appears to us to be without structure or meaning when seen in isolation. The scaling transforms it into geometric information, which is the type of information in which we find it easiest to recognise patterns. We saw that even in a small well-studied group of cases that multi-dimensional scaling helped us to understand the structure and complexity of the cases and that it suggested direction for further study.

But most interesting of all is to speculate upon the general validity of the nearest neighbour rule, for it is this rule which takes us from the domain of analysis to that of decision.

For the sake of argument, suppose that the general validity of the rule could be verified, at least for particular categories of case law.[50] If parties to a dispute could agree upon the facts of the dispute, the result would be a mechanical means of deciding cases.

This possibility of machine made decisions is a disconcerting one, stirring the worst of our Orwellian nightmares. But suppose that we restrict our attention to some area of the law where high costs or manpower requirements result in disputes in that area being outside the normal processes of the court.[51] If such disputes could be settled by a machine using the nearest neighbour rule (or some other algorithm) and if adequate safeguards were allowed for human review when required, is the prospect really so frightening?

These speculations, although taking us outside the scope of this paper, may serve to illustrate the ease with which the early jurimetricians were led into making extravagant claims for the future of the subject. I hope that they also serve to lead others to become interested in the study of jurimetrics.

ALAN L. TYREE.

---

50. The rule must be rephrased as a testable statistical hypothesis.
51. Small claims may be such an area of law.