Volume 2
Issue 1 *Computer/Law Journal - 1980*

Article 29

1980

Mathematical Models For Legal Prediction, 2 Computer L.J. 829 (1980)

R. Keown

MATHEMATICAL MODELS FOR LEGAL PREDICTION

by R. KEOWN*

INTRODUCTION

This article discusses some ideas concerning the prediction of judicial decisions by means of mathematical models. In a recent article, Mackaay and Robillard developed the "nearest neighbour rule" and collected a number of references to schemes predicting the outcomes of cases presented to a judicial body for decision.[1] Their paper credits Kort with the initial effort in this field.[2] Besides their own method of nearest neighbours, the two authors discuss procedures of McCarthy employing ideas from the theory of "artificial intelligence" to develop logical methods of case analysis and prediction,[3] various approaches of Lawlor concerning linear prediction schemes,[4] and review work of others—a description of whose efforts will not be attempted here.

This article begins with the method of linear models, goes on to that of catastrophic models, and concludes with the scheme of nearest neighbours. Unfortunately, the concepts required for a full understanding of catastrophe theory do not form a part of the standard equipment of most professional mathematicians, much less that of most practicing attorneys, so that this article must be directed primarily toward exhibiting the general flavor of the subject, rather

---

* B.S. physics 1946, University of Texas; Ph.D. mathematics 1950, Massachusetts Institute of Technology; candidate for J.D. 1981, University of Arkansas. Since 1967, Dr. Keown has been a professor of mathematics at the University of Arkansas, Fayetteville, Arkansas.

1. Mackaay & Robillard, *Predicting Judicial Decisions: The Nearest Neighbour Rule and Visual Representation of Case Patterns*, 3 DATENVERARBEITUNG IM RECHT 47 (1974).

2. Kort, *Predicting the Supreme Court Decisions Mathematically: a Qualitative Analysis of the "Right-to-Counsel" Cases*, 51 AM. POL. SCI. REV. 1 (1957).

3. McCarthy, *Reflections on Taxman: An Experiment in Artificial Intelligence and Legal Reasoning*, 90 HARV. L. REV. 837 (1977).

4. Lawlor, *What Computers Can Do: Analysis and Prediction of Judicial Decisions*, 49 A.B.A.J. 337 (1963).

than toward a penetrating analysis. In this regard, Zeeman gives one of the most satisfactory presentations of the basic ideas for non-mathematical readers.[5] Recently *Behavioral Science* devoted an entire issue to applications of catastrophe theory in the behavioral and life sciences, including a critique by two well-known detractors, Sussman and Zahler.[6] With these difficulties in view, it may be worthwhile to note that since the technique of linear programming has been widely introduced during the past twenty-five years, the notion of a general linear equation is understood by a much wider audience than ever before. Consequently, an effort will be made to describe some of the fundamental objects of catastrophe theory through the agency of linear models.

Before beginning with these models, it is necessary to introduce a few geometric concepts in an algebraic setting. Doubtless, these will be too elementary for some and too distracting for others. Nevertheless, some sort of an introduction is required to this complex of ideas and notation. Two of the basic notions of geometric analysis are that of a *plane* P consisting of all the ordered pairs (x, y) where x and y are real numbers and that of a *line* L, which is the set "{(x, y): ax + by + c = 0 for some numbers, a, b, and c}." This last "{ : }" is standard mathematical notation for a set S = {x: p(x)} consisting of all objects x where x is an object with property p. For example, J = {x: x is a federal judge} denotes the set of all federal judges; D = {x: x is a defendant in a California court during 1978} is the set of all defendants which appeared in California courts during 1978; and S = {x: x is a rule in the Income Tax Regulations of 1977} is the set of all rules which appear in the Income Tax Regulations of 1977.

The concepts of plane and line are *geometrical* in that they are geometric ideas historically as well as in the sense that meaningful pictures can be drawn of them. These are as follows:

5. E. ZIEMAN, CATASTROPHE THEORY: SELECTED PAPERS (1972-1977) (1977).

6. Sussman & Zahler, *A Critique of Applied Catastrophe Theory in the Behavioral Sciences*, 23 BEHAV. SCI. 383 (1979).

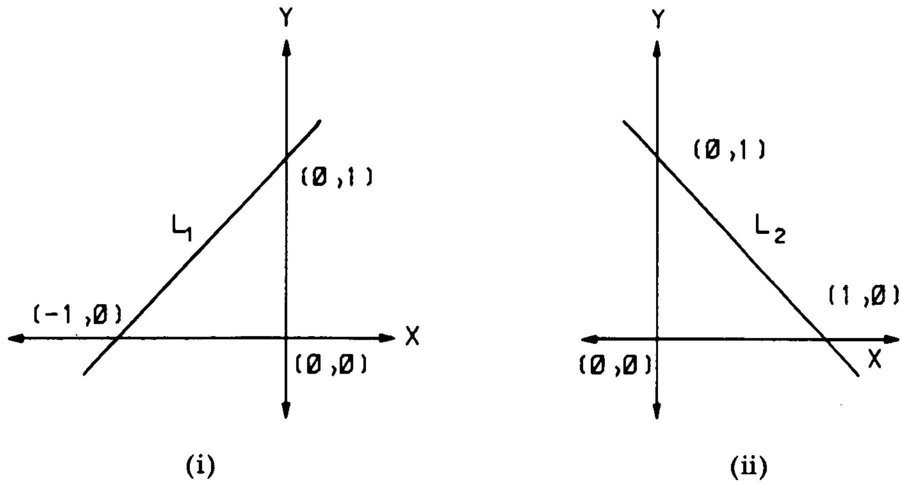

FIGURE 1

Figure 1(i) is a "picture" of the so-called x, y-plane containing the line L_1 of all points with coordinates (x, y) which satisfy the equation $x - y + 1 = 0$, while Figure 1(ii) is the "same plane" containing the line L_2. Intuitively the plane is a *two-dimensional* object while the embedded line is a *one-dimensional* object.

Such intuitive concepts of dimension can be given forbiddingly mathematical definitions and, based on these, one can show that most of the ordinary ideas of dimension are satisfied. Mathematicians refer to the space E_3 in which we live as *three-dimensional Euclidean space*. They represent E_3 both geometrically and algebraically as the set of all triplets (x, y, z) where x, y, and z are real numbers, or, in symbols, by $E_3 = \{(x, y, z): x, y, z \text{ are real numbers}\}$. It may be pictured in the form:

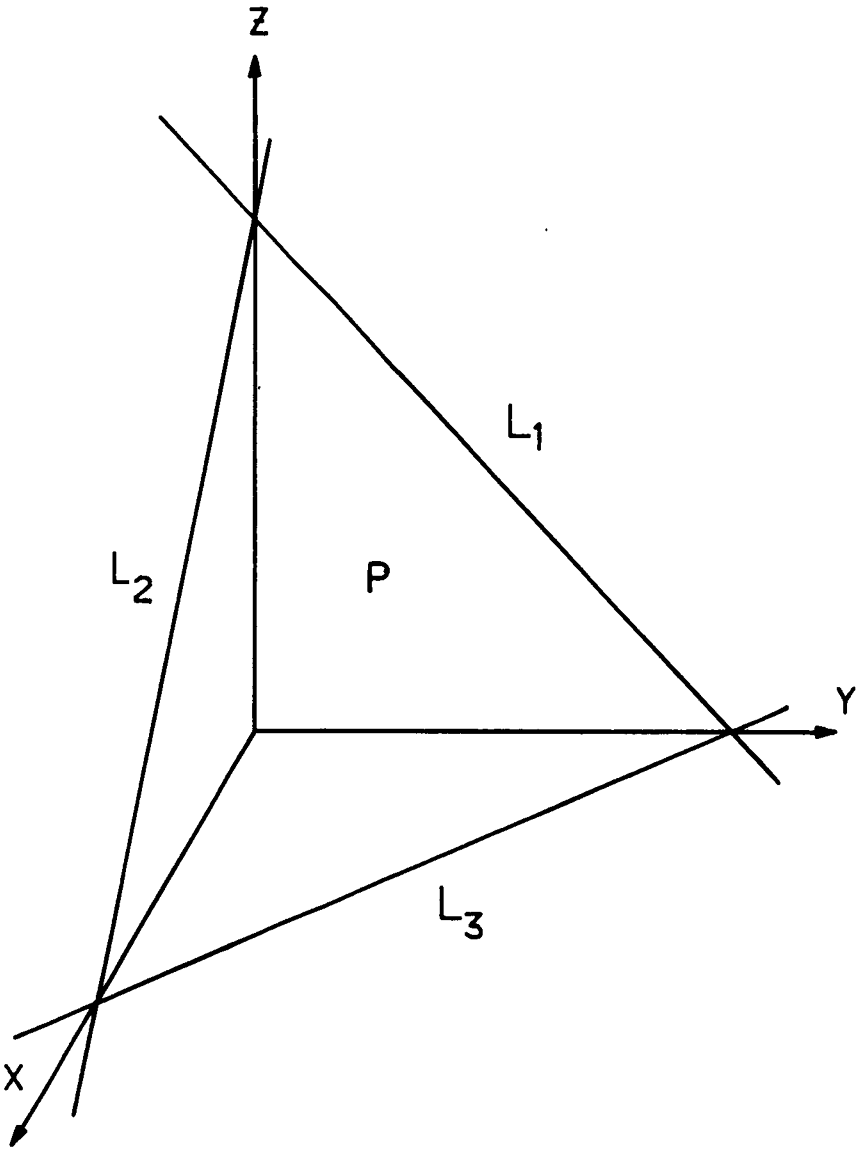

FIGURE 2

$L_1 = \{(0, y, z): 2y + 3z = 6\}$

$L_2 = \{(x, 0, z): 6x + 3z = 6\}$

$L_3 = \{(x, y, 0): 6x + 2y = 6\}$

$P = \{(x, y, z): 6x + 2y + 3z = 6\}$

Here L_1, L_2, L_3, and P denote three lines and one plane contained or embedded in the three-dimensional space E_3.

Although the physical intuition of most individuals declines sharply past three-dimensions, physicists have found many applications of the idea of four-dimensional space consisting of all four-tuples (x_1, x_2, x_3, x_4) of real numbers. Such four-dimensional spaces appear regularly in the theory of relativity.

Actually, many situations arise in which an n-tuple $(x_1, x_2, \ldots, x_n)$ of real numbers makes perfectly good sense. For example, suppose that one examines all the cases which have been decided in the Connecticut Supreme Court concerning zoning amendment appeals. A study of these cases may reveal a number of common issues. If a given issue, say *change in the character of the neighborhood* denoted by x_1, appears in the case, then x_1 takes the value of 1 while if it does not appear x_1 takes the value 0. A partial list of issues and their labels x_1, x_2, etc. might consist of

x_1 change in the character of the neighborhood
x_2 the new use is not needed
x_3 the new use is compatible from an economic standpoint
x_4 the new zone change is detrimental to the neighborhood
x_5 the character of the neighborhood supports the change
(along with perhaps a hundred and forty-five others).

A lawyer might use five-tuplets to describe such patterns, *i.e.*, the five-tuplet (0, 0, 1, 1, 1) could denote the situation where, $x_1 = 0$, denotes there has been no change in the character of the neighborhood, $x_2 = 0$, the new use is needed, $x_3 = 1$, the new use is compatible from an economic perspective; $x_4 = 1$, the zone change is detrimental to the neighborhood; $x_5 = 1$, the character of the neighborhood supports the new use. If there are 150 issues, then a 150-tuplet of the general form

$$(x_1, x_2, \ldots, x_{149}, x_{150})$$

describes a general fact pattern. Mathematicians call a space made up of n-tuplets $(x_1, x_2, \ldots, x_n)$ an *n-dimensional space*. Lines, planes, and three-space are examples of geometric objects which mathematicians call *manifolds*. A line L constitutes a *linear submanifold* of a plane P which contains it, while P itself may be a linear submanifold of a three-dimensional manifold E_3 (ordinary Euclidean space). The list of manifolds includes the n-dimensional spaces mentioned above which are referred to as *n-dimensional manifolds*.

The word *linear* is used in this connection because of the nature of the algebraic equations used to describe the subspace. Equations of the form

$$ax + by + c = 0, \text{ and}$$
$$ax + by + cz + d = 0$$

are called *linear equations*; the first is linear in the two *unknowns* x and y, and the second in the three unknowns x, y and z. An expression of the form

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n + c = 0$$

is a linear equation in n unknowns $x_1, x_2, \ldots, x_n$. A line L is said to be *linear*, because the coordinates (x, y) of a point of L satisfy a linear equation in two unknowns, while a plane P is said to be linear because the coordinates of a point (x, y, z) of P satisfy a linear equation in three unknowns. The set

$$H_n = \{(x_1, x_2, \ldots, x_n): a_1x_1 + a_2x_2 + \ldots + a_nx_n = c\}$$

is called an affine *hyperplane* of the space E_n. A hyperplane is a linear submanifold of E_n since the coordinates of any point H_n satisfy a linear equation in n unknowns, $x_1, x_2, \ldots, x_n$.

The number of data points required to determine a line $L = \{(x, y): ax + by + c = 0\}$ is two, as can be seen with a piece of paper and a straight edge. Frequently people wish to represent a set of data containing more than two points by the *most suitable* or *best straight line* as illustrated in the following drawing.

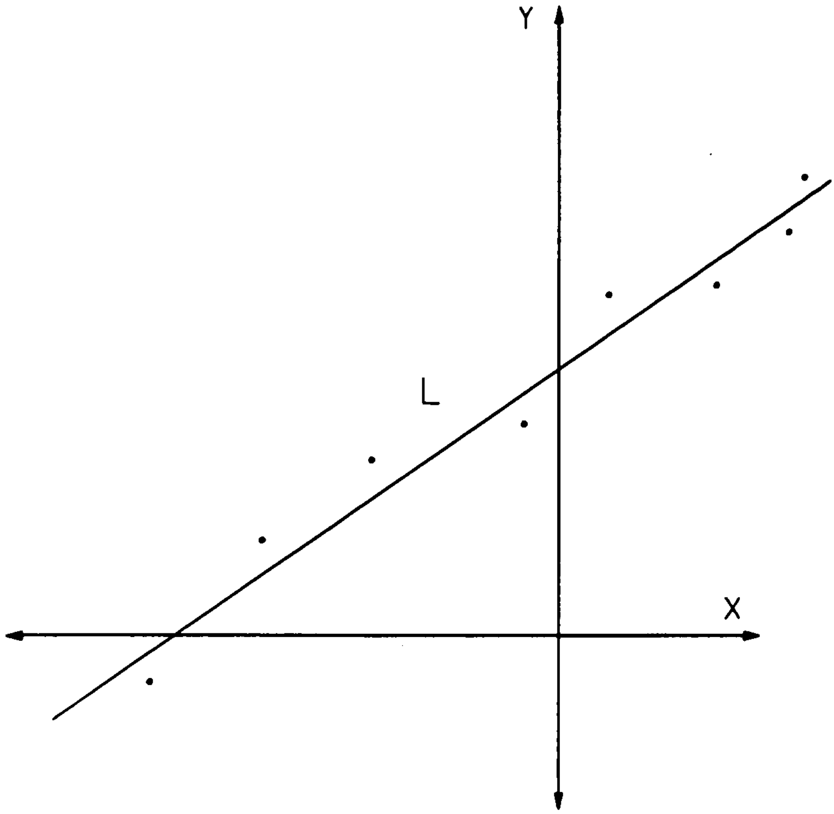

FIGURE 3

There are many mathematical techniques for determining such a line with the "method of least squares" being one of the most popular, while drawing the line by eye is generally satisfactory for a small number of points.

All the manifolds mentioned above are of infinite extent, but there are also many manifolds of finite extent. Perhaps the most familiar of these is the common sphere pictured below.

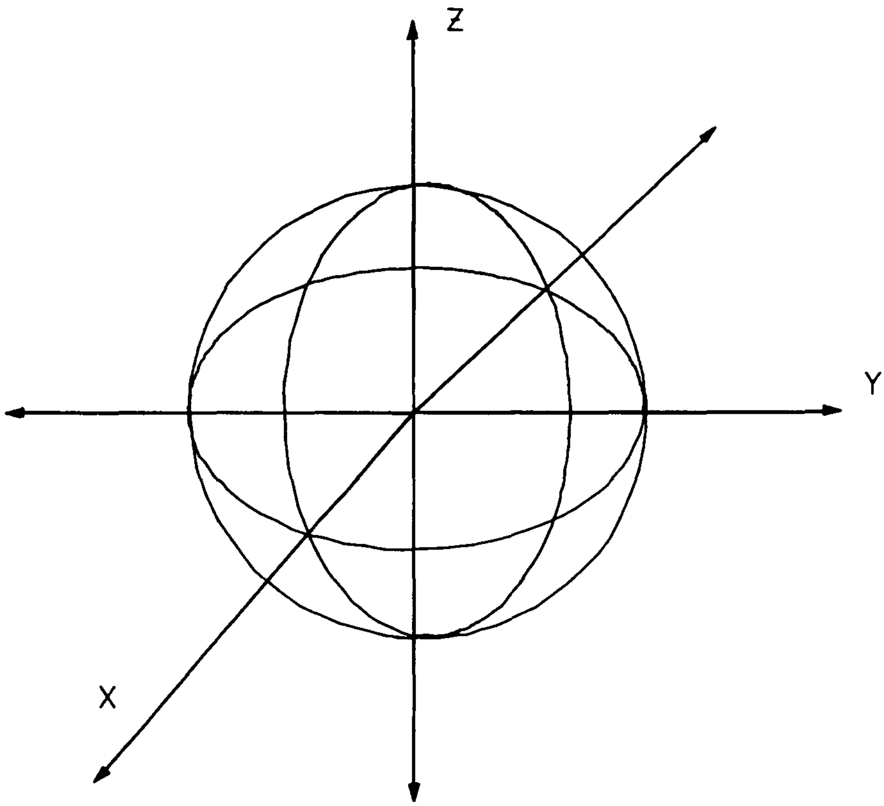

FIGURE 4

The algebraic equation satisfied by the coordinates (x, y, z) of each point of a sphere S of radius 2 is $x^2 + y^2 + z^2 = 4$. Except for small perturbations due to mountains, valleys, and volcanoes, the earth is almost a sphere, although men for generations thought it was a plane. This property of *looking like a plane locally* is the fundamental characteristic of a two-dimensional manifold.

An ellipsoid is a sort of distorted sphere having the general form of a football as shown below.

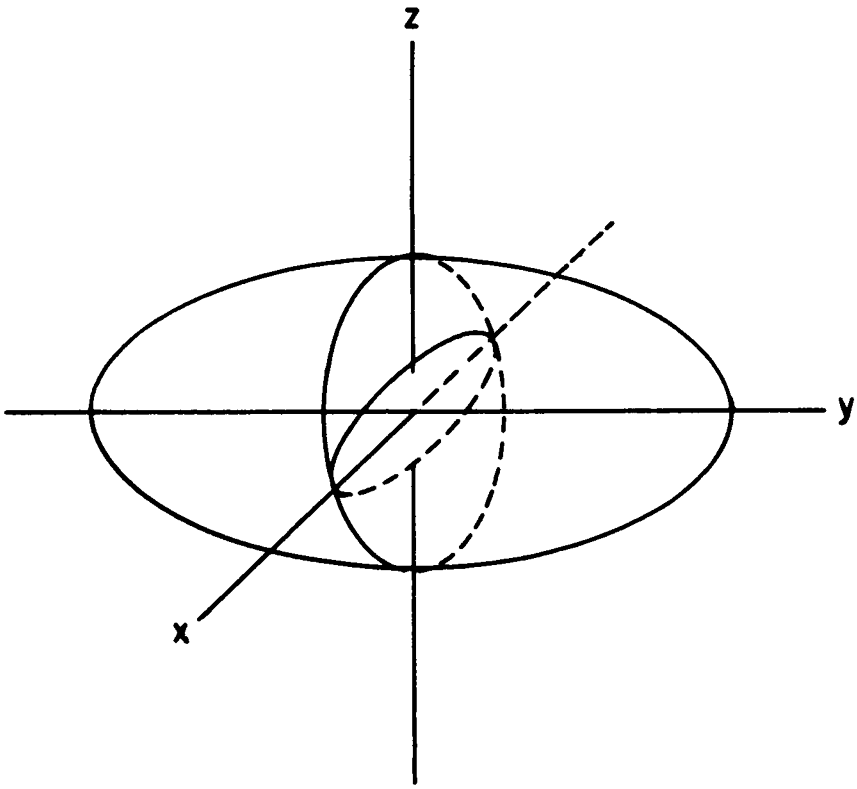

FIGURE 5

One can distort the sphere more seriously by attaching a "handle" to form another two-dimensional manifold with the shape of a donut, or add two handles to form a donut with two holes. This process generates an infinite family of two-dimensional manifolds: the sphere, the sphere with one handle, the sphere with two handles, . . . , the sphere with n handles, and so forth and so on. A picture of a fairly general two-dimensional manifold is given below.

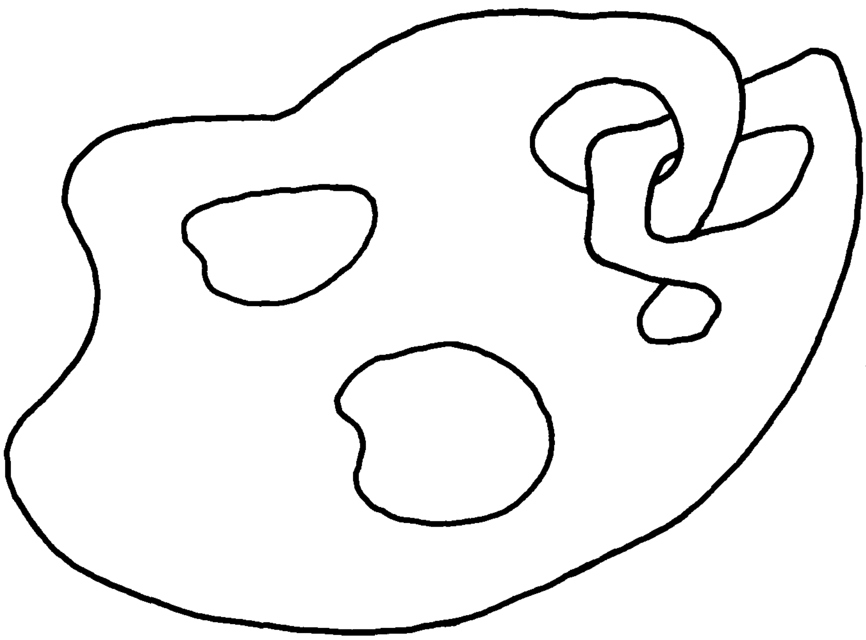

FIGURE 6

Before leaving the concepts of lines, planes, and manifolds, it is convenient to introduce the notion of rate of change of a quantity. If it costs $5 to type one letter in the usual law office, the cost of n letters is given by the formula, c = 5n, which determines the graph of Figure 7,

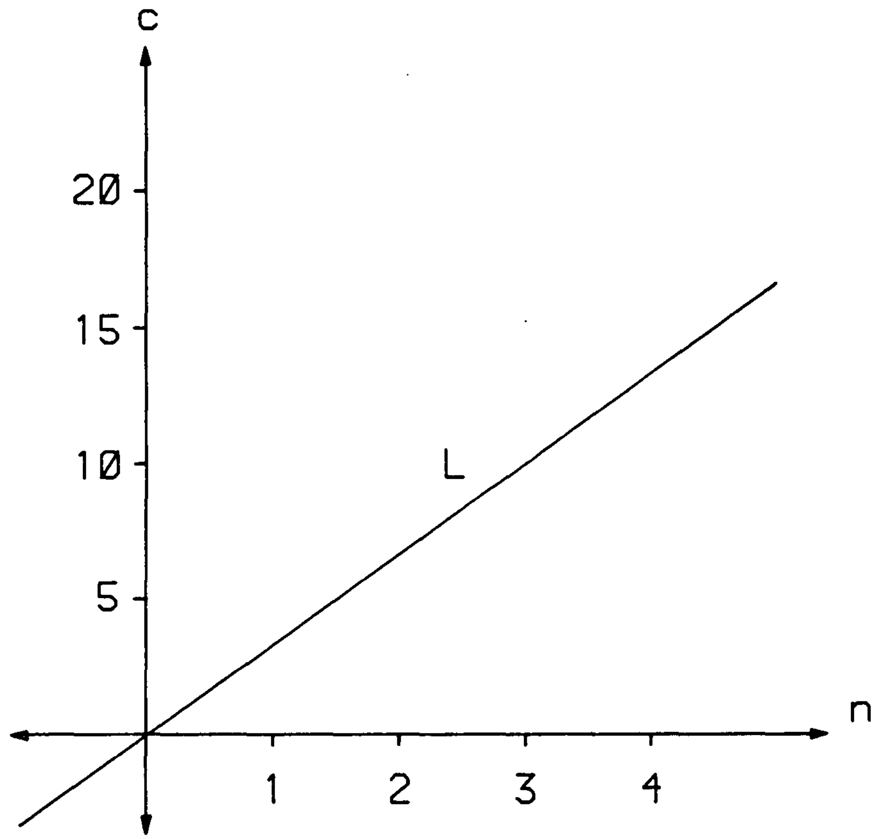

FIGURE 7

with the property that when the number of letters n increases by 1 unit, then the cost c increases by 5 units. Disregarding the fact that the symbol n is physically meaningful only when it denotes an integer or whole number, it is a geometric fact that c increases by 5 units on the graph when n increases by 1 unit no matter where one starts, whether at an integer or not. Mathematicians say that the *rate of change of* c *with respect to* n is 5 and denote this rate by the symbol dc/dn. If the monthly overhead of running a law office depends on five independent factors denoted by x_1, x_2, x_3, x_4, and x_5 with the respective unit costs (or rates of change) p_1, p_2, p_3, p_4 and p_5, then the total overhead cost is given by

$$c = p_1x_1 + p_2x_2 + p_3x_3 + p_4x_4 + p_5x_5.$$

The graph of this formula can be properly drawn only in a physical space of six or more dimensions, but the notion of rate of change remains the same, namely, it is the increase in c with respect to a unit increase in *any one* of the five quantities x_1, x_2, x_3, x_4 and x_5 and the notation is still much the same, *i.e.*, the five rates of change are denoted, respectively, by the symbols $dc/dx_1 = p_1$, $dc/dx_2 = p_2$, $dc/dx_3 = p_3$, $dc/dx_4 = p_4$, and $dc/dx_5 = p_5$. These rates of change are all constant because the previous functional descriptions are all in terms of linear models useful in many cases but somewhat deficient in others. To illustrate the distinction between linear and nonlinear models, consider a variable u depending upon a variable n through a function relation with the graph sketched in Figure 8.

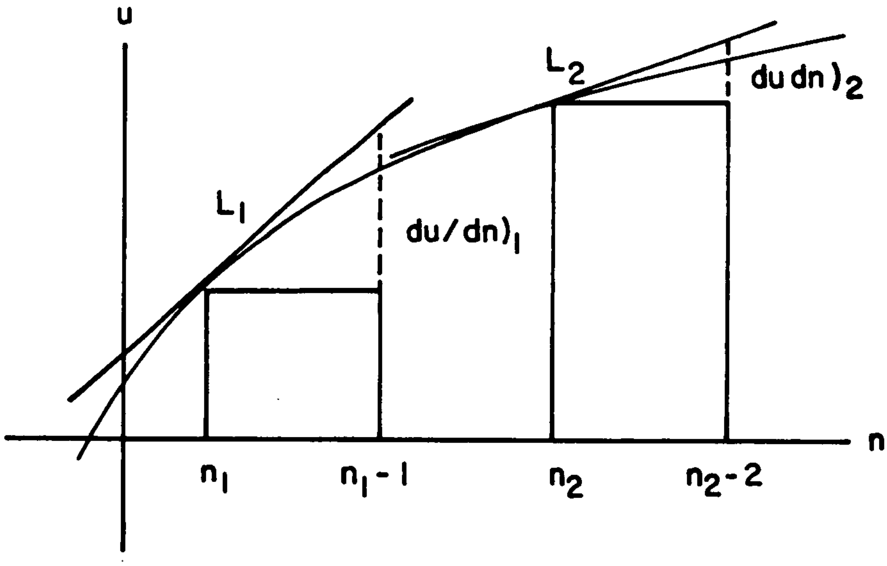

FIGURE 8

When the functional relationship between u and n is nonlinear as above, then it becomes necessary to speak of the *instantaneous rate of change* $du/dn)_1$ of the variable u with respect to the variable n when, for example, $n = n_1$.

Mathematicians invented a method for computing this instantaneous rate of change some two hundred years ago, but a rather good estimate can be had by merely drawing the line L_1 of Figure 8 which is the *best linear* approximation to the curve or graph at the point (n_1, u_1). If the equation of the line is $u = m_1n + b_1$, then the rate of change along L_1 is the number m_1 (sometimes called the *slope* of

L_1). The slope m_1 is equal, by definition, to the instantaneous rate of change of u with respect to n at the point (n_1, u_1), and also equal to the length of the dotted vertical line at $n_1 + 1$. A similar "best straight line or linear approximation" to the curve at (n_2, u_2) is given by the line L_2 whose equation is $u = m_2n + b_2$ so that the instantaneous rate of change $du/dn)_2$ equals m_2, numerically the same as the length of the dotted vertical line at $n_2 + 1$. The instantaneous rate of change of u with respect to n is clearly larger at (n_1, u_1) that it is at (n_2, u_2).

I. LINEAR MODELS

The method of linear models relates to the concept of "best linear fit" as illustrated in Figures 3 and 8, but in most situations researchers generally apply the method to n-dimensional hyperplanes in (n + 1)-dimensional space. There the x-axis of Figure 3 becomes an n-dimensional "coordinate hyperplane" and the vertical y-axis remains as shown. The data to be fit consists of a collection of P_1, P_2, . . . of "points" of the form $(y, x_1, x_2, \ldots, x_n)$ requiring an additional subscript to distinguish one point from another as indicated below:

$$P_1 = (y_1, x_{11}, x_{12}, \ldots, x_{1n})$$
$$P_2 = (y_2, x_{21}, x_{22}, \ldots, x_{2n})$$
$$P_3 = (y_3, x_{31}, x_{32}, \ldots, x_{3n})$$

One can imagine $P_1, P_2, \ldots, P_k$ points scattered about an (n + 1)-dimensional data space E_{n+1} so that the problem is one of determining the *best linear approximation*, as in the case of the line in Figure 3.

As previously mentioned, mathematicians have developed various techniques for finding such a "best hyperplane", but the most popular method is probably that of *least squares*. This method selects the desired hyperplane so as to minimize the sum of the squares of the distances from the hyperplane to the data points. Many computer programs are available to perform these kinds of calculations. All lead to a linear expression for y in terms of the variables $x_1, x_2, \ldots, x_n$ of the form

$$y = a_1x_1 + a_2x_2 + \ldots + a_nx_n + a_{n+1}$$

where the a's have been determined so as to minimize the sum of the squares. Generally, statisticians prefer to have many more data points $P_1, P_2, \ldots, P_k$ than variables $x_1, x_2, \ldots, x_n$ since this allows them to make informed estimates about the accuracy of the determination of the a's, but the present article does not explore this point.

The expression for y obtained above is called a *linear predictor*.

The use of linear predictors of the outcome of court decisions has been advocated for at least twenty years. Some of the first innovaters were Fisher,[7] Schubert,[8] and Baade.[9] Perhaps the most active person in the field has been Reed Lawlor.[10] The person who has achieved the widest acclaim is undoubtedly Haar who, with Sawyer and Cummings, made a year long study of zoning amendment cases brought before the Supreme Court of Connecticut. The model of Haar is described in detail in the *American Bar Foundation Research Journal*[11] and has recently enjoyed great success. According to the *Commercial Law Journal*, the model of Haar gave the correct predictions in ninety-nine percent of over one thousand cases selected from a variety of states.[12] Even without this surprising record, it is well worth considering the work of Haar and his associates as a tutorial in methodology.

Haar, Sawyer and Cumming formed a team of two lawyers and one statistician to make a statistical study with the aid of a computer of the zoning amendment cases decided by the Connecticut Supreme Court. The two lawyers identified a collection of 167 issues from seventy-nine cases heard by the Connecticut court over a period of roughly twenty years. Only forty of the issues proved significant when statistical tests were run.[13] Since seventy-nine cases supply too little data for a good linear fit in a 41-dimensional space, the investigators used a method of analysis which groups the variables having similar effects on the outcome into subsets, called *scales*.[14] As a result of this procedure, called *factor analysis*, the variables were grouped to form the eleven scales listed below:

Scale 1 Compatibility Indicated by Change in the Character of the Neighborhood
Scale 2 Use not needed
Scale 3 Adequate Physical Planning
Scale 4 Public Interest Planning and Zoning Techniques
Scale 5 Compatibility from an Economic Perspective
Scale 6 Zone Change Detrimental

---

7. *See* Fisher, *The Mathematical Analysis of Supreme Court Decisions: The Use and Abuse of Quantitative Methods*, 52 AM. POL. SCI. REV. 321 (1958).

8. *See* JUDICIAL DECISION-MAKING (G. Schubert ed. 1963).

9. *See* JURIMETRICS (M. Baade ed. 1963).

10. *See* Lawlor, *Foundations of Logical Legal Decisions Making*, M.U.L.L. 98 (1963).

11. Haar, Sawyer, & Cummings, *Computer Power and Legal Reasoning: A Case Study of Judicial Decision Prediction in Zoning Amendment Cases*, 1977 AM. B. FOUND. RES. J. 651 (1977) [hereinafter cited as Haar].

12. 85 COM. L.J. 270 (1980).

13. Haar, *supra* note 11, at 711.

14. *Id.* at 712.

Scale 7 Physical Services Inadequate
Scale 8 Compatibility Indicated by Large Uniform Blocks
Scale 9 Good Planning Practices
Scale 10 Character of Area Supports Change
Scale 11 Large-Area Zoning[15]

The investigators retained two of the original variables in the regression analysis (statistical language for linear fits and linear prediction models):

011 The Court of Common Pleas Approved/Denied the Zone Change
012 The Zoning Authority Denied/Approved the Zone Change

They omitted these two variables from the factor analysis, since they had previously decided to include them in their linear predictor.[16] The research of Haar and his colleagues produced three models of the general form

$$P = A + C_1V_1 + \ldots + C_kV_k$$

where $A = 0.56523$ and the values of the coefficients $C_1, \ldots, C_k$ are listed in Table 1 below.[17]

TABLE 1

COEFFICIENTS OF THE LINEAR MODELS

Variable	Basic Model	Model with Scale 9 added	Model without Scale 4
Scale 4	.05769	.05518	—
Scale 6	–.19320	–.18397	–.19139
Scale 7	–.25548	–.24747	–.25450
Scale 8	.21506	.20939	.21336
Scale 9	—	.04473	—
Scale 10	.05265	.05629	.06746
Scale 11	.11884	.10477	.12014
Var. 011	.16121	.16510	.18324
Var. 012	–.55106	–.53905	–.59198

The linear predictor for the Basic Model can be written $P = 0.56523 + 0.05769S_4 - 0.1932\ S_6 - 0.25548\ S_7 + 0.21506\ S_8 + 0.05265\ S_{10} + 0.11884\ S_{11} + 0.165\ V_{011} - 0.55106\ V_{012}$. The variables $S_6, \ldots, V_{012}$ assume only the values 0 and 1 so that those terms with plus signs designate issues primarily for plaintiff and those with minus signs those issues primarily for defendant.

Based upon the facts of a particular case, *e.g.*, the Court of Common Pleas has approved, so that $V_{011} = 1$ or has disapproved, so that $V_{011} = 0$; the Zoning Authority denied the zone change, so that V_{011}

15. *Id.* at 713.
16. *Id.* at 712.
17. *Id.* at 716.

= 1 or will have granted the change so that $V_{011} = 0$, and so forth for the remaining variables, one obtains a numerical value for P. Under the conventional notion that plaintiff wins with a preponderance of the evidence in his favor, the linear model predicts victory for plaintiff whenever the entries for the given fact pattern produce values greater than 0.5 for P.

Haar and his associates reported the interaction of their linear predictor with predictions by conventional legal analysis.[18] Haar asserts that this sort of approach leads to a better organized attack on a legal problem, since preparation for computer analysis requires a very systematic examination of the cases.[19] Furthermore, he claims that computer modelling displays the *ratio decidendi*[20] of the case as a relationship between the facts and outcome, which can be mathematically expressed to the extent that the lawyer can ascertain precisely what facts were before the court.[21] Moreover, the close reading required for computer analysis of the case improves the determination of the key factors in a case-by-case evaluation.[22] Finally, this type of organization eases the task of a newcomer in the selected area of law by the systematization of the material.[23]

II. Catastrophic Models

Researchers have developed two basic approaches to the theory of catastrophes which may be called, in the spirit of physics, the *theoretical* approach and the *phenomenological* one. Rene Thom created the theoretical while Christopher Zeeman created the *phenomenological*. The next several pages discuss Thom's method on which the models of Zeeman are based. Readers wishing to skip the theoretical discussion may proceed to subsection II.B *infra* which introduces the ideas of Zeeman.

A. *Thoms Theoretical Approach*

Thom invented his theory by means of ideas apparently borrowed from classical mechanics, since the rate of change of a key variable depends on a potential, of which the most familiar are those created by gravity. A skier standing at the crest of a ridge must exercise great care merely to maintain himself there while, to the contrary, one located in a valley must exert great effort to get out. The

18. *Id.* at 742-51.
19. *Id.* at 745-46.
20. *Id.* at 746.
21. *Id.*
22. *Id.*
23. *Id.*

first skier is said to be in a position of unstable equilibrium and the second in a position of stable equilibrium. To fix the ideas, consider a cross-section of the country side assuming the form shown in Figure 9.

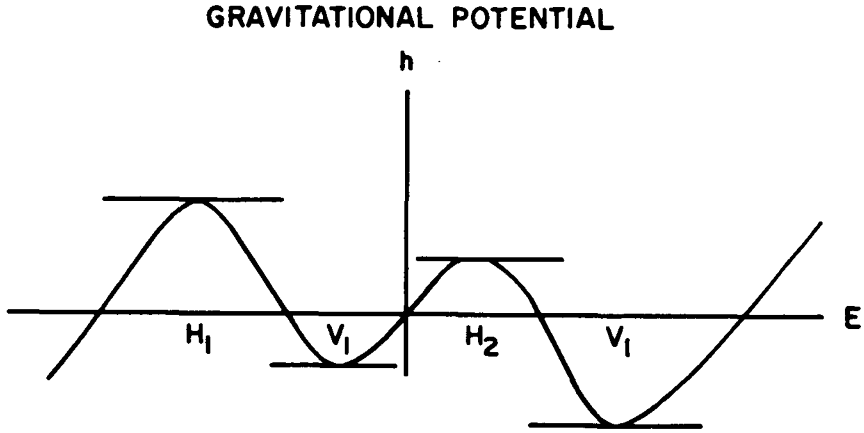

FIGURE 9

GRAVITATIONAL POTENTIAL

Here the vertical variable h represents height above sea level while the horizontal variable E represents a line in an East-West direction. The position determined by H_1 denotes the crest of a hill on which a particle, say a marble, would remain balanced if it were not displaced "at all" from its position. Physicists and catastrophe theorists refer to such a point as one of *unstable equilibrium*. The position V_1 depicts the bottom of a valley at which the marble would return even if displaced a small distance. Theorists call this a point of *stable equilibrium*. H_2 and V_2 locate additional unstable and stable equilibrium points, respectively. Figure 10 sketches a more realistic, three-dimensional presentation of such a situation.

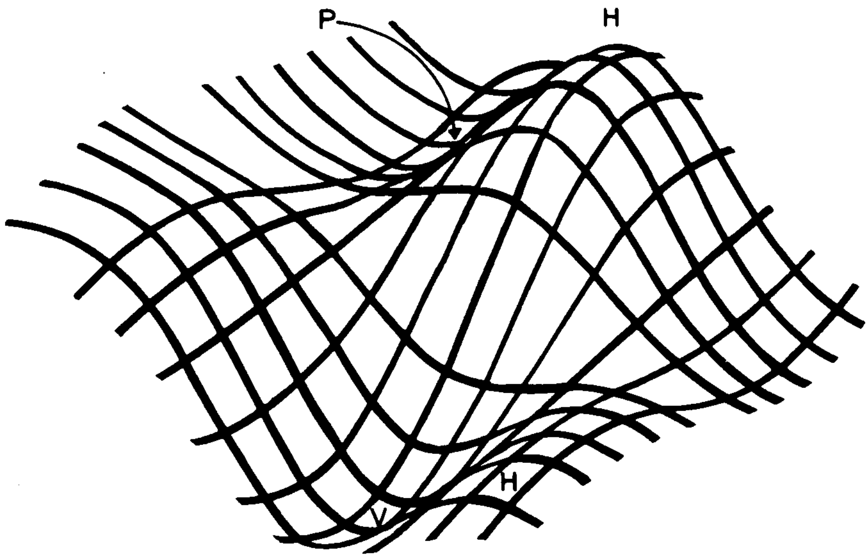

FIGURE 10

This figure illustrates three kinds of *critical points* which are naturally related to the stable and unstable equilibrium points of Figure 9. The reader should not allow the picture to suggest these exhaust all possible kinds of critical points since they form but a small selection. The points H, V, and P represent critical points which could be called *hills, valleys,* and *passes*. Only the valleys are stable equilibrium points; the other two are unstable.

These figures illustrating gravitational potentials display certain aspects of catastrophe theory whose potentials depend on auxiliary variables called *control parameters*. One of the simplest potentials occurring in the theory is

$$V(a, x) = x^3/3 - ax$$

where attention focuses on the dependence of V(a, x) on the control parameter a. This dependence is graphically illustrated in Figure 11.

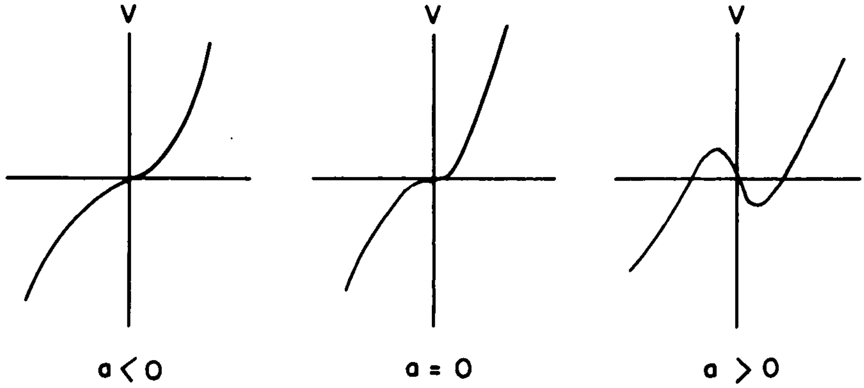

FIGURE 11

CATASTROPHIC POTENTIALS

Geometrically, these graphs show no equilibrium point when a is negative, one unstable equilibrium point at the origin when a equals 0, and two equilibrium point at the origin when a is positive, namely, an unstable equilibrium point at x = $-a$ and a stable equilibrium point when x = a. These equilibrium points may be determined from the rate of change dV/dx, which has the form (the negative is used in catastrophe theory)

$$-dV/dx = -x^2 + a = -(x^2 - a).$$

They are those points determined by values of x for which the rate of change dV/dx is zero, *i.e.*, that satisfy the equation

$$x^2 - a = 0$$

This gives the x values listed for points where the tangent line to the curve is horizontal. As mentioned previously, catastrophe theory revolves around the dependence of the equilibrium points on the control variable a. It proves useful to depict this dependence graphically, partly to introduce some additional terminology.

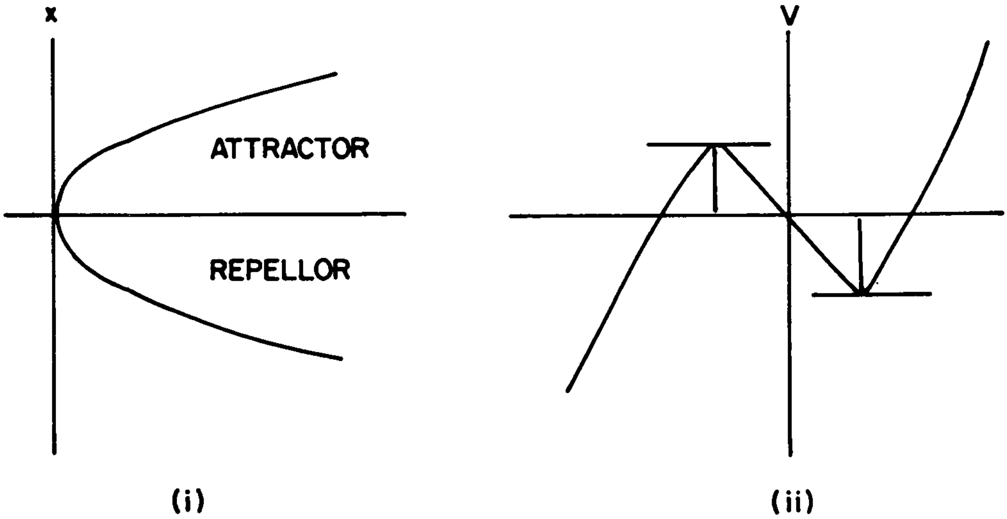

FIGURE 12

FOLD CATASTROPHE

The relation $x^2 - a = 0$ generates two functions, $x = a$ and $x = -a$. The graph of the first gives the upper branch labeled *attractor* and that of the second gives the lower branch labeled *repellor* in Figure 12(i). If $a = 1$, the potential $V(1, x)$ has an unstable equilibrium point at $x = 1$ as shown in Figure 12(ii). This choice, $a = 1$, illustrates the fact that x-values on the upper (attractor) branch of Figure 12(i) correspond to stable equilibrium points while x-values on the lower (repellor) branch correspond to unstable equilibrium points.

Catastrophe theorists call the graph 12(i) the *behavior manifold* of the system with potential $V(a, x)$. In this particular instance, the behavior manifold constitutes a one-dimensional manifold in the two-dimensional manifold of all pairs (a, x) where x is the *response* and a is the *control variable*. The point $(0, 0)$ on the behavior manifold is a *catastrophe point* which implies that a small change along the attractor curve will produce a point of stable equilibrium, while a small change along the repellor curve will produce a point of unstable equilibrium.

The original investigators based their analysis on equations of the form

(1) $$dx/dt = -dV/dx$$

which relate the rate of change of the response x with respect to t to the rate of change of the potential V with respect to x. They called the resulting process a *gradient system*. The potential $V(a, x) = x_{3/31} - ax$, depending on a single control parameter a, generates the simplest gradient system having a behavior manifold like that of Figure 12(i) with a single catastrophe point. The next simplest type of catastrophe has a potential with two control parameters, a and b, with the expression

(2) $$V(a, b, x) = x^4/4 - ax - bx^2/2.$$

This potential gives rise to the gradient system

(3) $$dx/dt = -(x^3 - a - bx).$$

For this system, the behavior manifold is the set of points in E_3 satisfying the relation

(4) $$x^3 - a - bx = 0.$$

This relation graphs into a two-dimensional submanifold of the three-dimensional manifold E_3 of all points of the form (a, b, x). The *control space* of this process is the a,b-plane consisting of all points with coordinates (a, b, 0). Figure 13 presents these details pictorially.

In our initial example, the catastrophe set consisted of a single point, but in the present it is the subset of the behavior manifold defined by

(5) $$C = \{(a, b, x): 3x^2 - 0\}.$$

This catastrophe set C is a one-dimensional curve contained in the two-dimensional behavior manifold. It forms the boundary of the cross-hatched area of Figure 13 and projects downwards into a curve B called the *bifurcation set* located in the control space. The behavior manifold is "pleated" into a sort of tuck over the region R bounded by the bifurcation set, *i.e.*, there are three distinct surfaces over the region R. The curve C separates the cross-hatched surface from an upper surface called *judgment for plaintiff* and a lower surface called *judgment for defendant*. This Judicial Model takes the vertical or x-axis as a *judicial axis* representing the outcome of a given fact pattern. Such a fact pattern is described by means of two auxiliary axes D and P, D denoting the evidence and argument for the defense and P denoting the evidence and argument for the plaintiff. The variables D and P are related to the control parameters a and b by two linear equations, such as

$$a = c_{11}D + c_{12}P + c_{13}$$
$$b = c^{21}D + c_{22}P + c_{23}.$$

We will spare the reader any further mathematical analysis and finish with a few remarks. It may be worthwhile to attempt to summarize in a single paragraph some of the astounding achievements of Thom. He showed that the exceptional or singular points of gradient systems, determined by equations such as Equation (1) above, could be classified into a finite number of "relatively simple" *types*, just as a lawyer might classify legal issues as procedural or substantive. These types give rise to a family of *standard models*, whose features depend, among other things, upon the number of control parameters in a standard potential. Not surprisingly, the complexity of the standard model increases as the number of control parameters in the standard potential increase. The present survey has been limited to the cases of one or two control parameters which lead, respectively, to the *fold catastrophe* of Figure 12 and the *cusp catastrophe* of Figure 13.

B. Zeeman's Phenomenological Approach

While the classification schemes of Thom depend upon an analysis of a gradient system involving a potential, the resulting models are presented in terms of *standard algebraic equations* that can be considered independently of their source. Zeeman advocates the use of these in the social sciences so that traditional analysis such as that given by the theory of linear models, *i.e.*, those depending on linear equations, can be supplemented with more complicated algebraic models depending on non-linear algebraic equations. Besides resistance by traditionalists, there exists

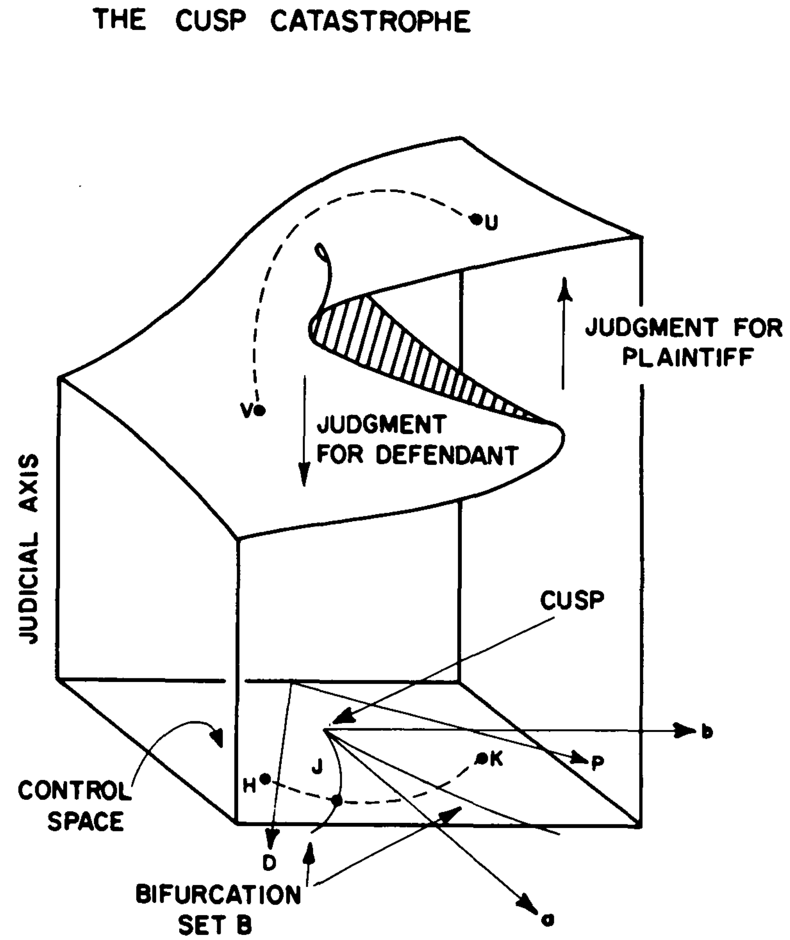

FIGURE 13

a serious problem involving the development of the required non-linear statistics. Cobb has made a promising beginning by returning to basics and incorporating a stochastic noise term into the standard potentials.[24]

Prior to the development of a rigorous statistics, there had been a flourishing evolution of *ad hoc* models in various social sciences

24. Cobb, *Stochastic Catastrophe Models and Multimodal Distributions*, 23 BEHAV. SCI. 360 (1978).

inspired by the pioneering work of Christopher Zeeman.[25] In the spirit of the physics motivating much of the earlier work, we refer to his methods as the *Zeemanian Process*. The following discussion depends substantially on insights cultivated by reading Zeeman's treatment of institutional disturbances.[26]

The fact pattern underlying a judicial decision comprises issues that may be classified either as (1) evidence and argument supporting the position of plaintiff denoted by the symbol P, or (2) evidence and argument supporting that of defendant denoted by D. In law, of course, who is plaintiff and who is defendant may depend on which party wins the race to the court house, rather than on the nature of the dispute involved.

In civil cases, plaintiff wins his case if the trier of fact, sometimes a judge and sometimes a jury, finds a preponderance of the evidence in his favor. Considering D and P as conflicting factors in a judicial process enjoying a suitably discontinuous behavor, one arrives by means of the Zeemanian process at:

Hypothesis I. The standard model of the judicial process is a cusp catastrophe with plaintiff's evidence and argument denoted by P and defendant's evidence and argument denoted by D as conflicting factors determining the outcome.

Figure 13 represents the standard model of the cusp catastrophe with the behavior manifold divided into parts—an upper part denoted as *judgment for the plaintiff* and a lower part denoted as *judgment for defendant*. Observe that in a V-shaped region loosely centered between the P and D axes these two pieces overlap with the upper surface joined to the lower by a cross-hatched area bounded by the *catastrophe curve*. The catastrophe curve projects downward into the *bifurcation set* which lies in the control space. Using nomenclature introduced in the theoretical discussion, both judgment for plaintiff and judgment for defendant are attractor surfaces which implies by the general theory that they are surfaces of stable equilibrium while the cross-hatched surface joining the two is a repellor, a surface of unstable equilibrium.

As intended, these terms indicate that, with certain exceptions to be discussed below, the *state of the case* is (defined by a point on the behavior manifold that indicates a victory either for plaintiff or for defendant at any given time). If the state of the case is represented by a point on judgment for plaintiff, then it tends to remain there, while if it is represented by a point on judgment for the de-

25. Probably one of the most useful sources of information on his techniques is E. ZEEMAN, note 5 *supra*.

26. *Id.* at 387.

fendant then by stable equilibrium it tends to remain there. A state of the case in which the outcome is uncertain corresponds to a point on the cross-hatched surface representing a situation of unstable equilibrium. Thus, continued introduction of evidence and further argument soon displaces the state either to judgment for plaintiff or judgment for defendant.

Since plaintiff normally addresses the court first, unless the case is dismissed for failure to state a cause of action, his attorney should be able to secure a position (D, P) in the control space, say K, determining a state belonging to judgment for the plaintiff. Starting with the situation at K, the defense attorney must try to drive the point on judgment for plaintiff onto the surface judgment for defendant. One method might be by producing evidence of perjury or of a damaging admission by plaintiff to reduce the value of P (evidence and argument for plaintiff) and move the state along the dotted path on the behavior manifold from U (above K) to V (above H). A second method might be to produce overwhelming evidence and argument for defendant so as to move from the point K of the control space to the point H along the dotted path in Figure 14, thereby reaching the same state V in judgment for defendant. This last path illustrates two phenomena of catastrophe theory.

The first is the concept of *delay* caused by overlapping of judgment for the plaintiff and judgment for defendant above the region R bounded by the bifurcation set. As one moves along the indicated path from K to H, the state of the case remains on the upper surface of judgment for plaintiff all the way over to the point on the catastrophe curve lying above J on the bifurcation set. As a result, the favorable state desired is delayed beyond the point where it normally should have occurred. The second is the phenomenon of *catastrophic jump*. When the motion reaches a point on the catastrophe curve above J, there is a catastrophic jump from judgment for the plaintiff to judgment for defendant. Observe that such a jump fails to occur if the first path of our argument is followed.

CATASTROPHIC PATH FROM K TO H

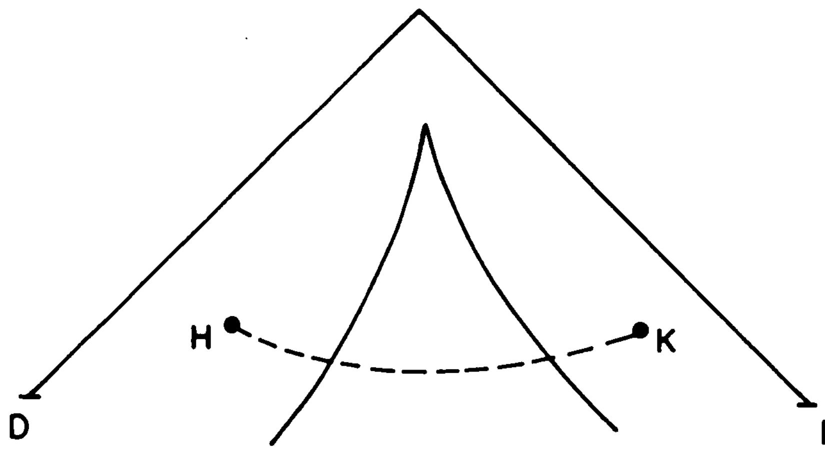

FIGURE 14

Small shifts occur in the structure of judicial catastrophic surfaces for a variety of reasons. For example, society contains many natural plaintiff-defendant pairs including mortgagees versus mortgagors, creditors versus debtors, and insureds versus insurors, along with hundreds more who are eternally trying to better their posture before the courts. As a consequence, many of them lobby for favorable legislation, litigate propitious rather than unpropitious cases, and include one-sided clauses favoring themselves in their contracts.

Hypothesis II. There is a continuing tendency for the judicial process to avoid the extremes of "judgment for the plaintiff" or "judgment for the defendant."

Zeeman considers this phenomenom as a sort of flow on the catastrophic surface representing a feedback from the parties to the courts, tending to forestall ultimate stability in the system. There are other influences on the judicial process created by such things as one jury being more objective than another, one lawyer being more effective than his opponent, and one judge having more judicial competence than a brother, with the result that in addition to feedback there is present a varying amount of what Zeeman calls "noise" in the process. He conceives such noise as forcing a particular case off the table locus on occasion. As a consequence, there may rise catastrophic transfers from the mere presence of noise in the system. Zeeman offers this observation as a hypothesis.

Hypothesis III. External events, or internal incidents within the judicial system may be represented as stochastic noise.

As this completes the description of catastrophic models, the discussion now turns to the method of nearest neighbours.

III. NEAREST NEIGHBOUR RULE

In their book on numerical taxonomy, Sneath and Sokal define their subject to be "the grouping by numerical methods of taxonomic units into taxa on the basis of their characteristic states."[27] These two scientists applied their methods primarily to biological sciences in which elaborate classification schemes have developed since the invention of taxonomy by Linnaeus in the eighteenth century. After the advent of large scale computing machinery, researchers introduced numerical methods into the subject. During the last quarter of a century, their procedures have spread from biology into many other areas, particularly medicine.

The emphasis on empirical analysis of data leads naturally to what may be called the "operational approach to taxonomy," by analogy with P. W. Bridgman's ideas.[28] In such a context "operationalism" implies that statements and hypotheses about a subject, law for instance, are subject to meaningful questions that can be tested by observation and experiment. In law, to determine whether Case A is more related to Case B than it is to Case C, clear definitions must be given of what is meant by "more related," that is, by what criteria "more or less relatedness" can be measured. This leads to the notion of assigning a set of characteristics to cases to divide them into groups or classes.

For example, Case A belongs to Tax Group 1 if, and only if, Case A has certain characteristics, *i.e.*, belongs to that collection of cases concerned with (1) income tax, (2) gross income, (3) prizes, (4) treasure trove, (5) possession . . . . Thus, a case belongs to a particular Tax Group if, and only if, it possesses all characters from a defining list. A group of cases so determined will be referred to as a *monothetic group*.

Under this format, a group of cases is defined by reference to a set P of properties which are both necessary and sufficient for mem-

27. P. SNEATH & R. SOKAL, NUMERICAL TAXONOMY 4 (1973).
28. *Id*. at 17.

bership in the class. It is possible, however, to define a group G in terms of a set

$$P = p_1, \ldots p_n$$

of properties in a somewhat less restrictive manner. Suppose we have a collection of cases such that

1) Each one has a large number of the properties in P.
2) Each p in P is possessed by large numbers of these individuals, and
3) No p in P is possessed by every individual in the aggregate.

According to condition 3, no single property p is necessary to membership in the collection; and nothing has been said to warrant or rule out the possibility that some p in P is sufficient for membership in the aggregate. A group of cases is *polythetic* if the first two conditions are fulfilled and is *fully polythetic* if condition 3 is also fulfilled. Wittgenstein has emphasized the importance of these ideas in ordinary language and especially in philosophy.[29] The common example in law of a line of cases tends to represent a *fully polythetic group* or *collection* of cases.

We introduce the following cases as illustrations of the ideas we have discussed.[30]

CASE A: *Cesarini v. United States*[31] concerned a tax deficiency declared by the Commissioner on the basis of some $4,500 found by the Cesarinis in a piano purchased in 1957. Having discovered the money in 1964, they reported it as ordinary income for that year. Later, the Cesarinis tried to amend their return, claiming there was either no tax due at all on the money was, at most, capital gains. The Circuit held the amount was ordinary income. Characters: 4, 6, 17, 23, 31, 34, 42, 49, 54, 55, 56, 61, 70, 71, 73, 74.

CASE B: *Old Colony Trust Co. v. Commissioner*[32] treats the claim of the government that W. M. Wood was in receipt of income whenever his company, American Woolen Company, paid his taxes of 1918 and 1919, that is, the payments made to the Internal Revenue Service by the company represented income to Mr. Wood. The Supreme Court agreed with the Commissioner. Characters: 10, 11, 13, 20, 28, 34, 48, 50, 63, 64, 68.

CASE C: *Commissioner v. Glenshaw Glass Co.*[33] raised the issue of whether money received as exemplary damages for fraud or

---

29. *Id.* at 21.
30. *See* Table 2 *infra* for a list of key characters for income tax cases.
31. 396 F. Supp. 3 (N.D. Ohio 1969), *aff'd per curiam*, 428 F.2d 812 (6th Cir. 1979).
32. 279 U.S. 716 (1929).
33. 348 U.S. 426 (1955).

for treble-damages in an antitrust recovery must be reported as gross income. The Supreme Court held these to be taxable income. Characters: 2, 6, 16, 25, 27, 31, 33, 34, 40, 56, 57, 68.

CASE D: *Chandler v. Commissioner*[34] considered whether or not the furnishing of a house to a principal stockholder by a corporation resulted in income to the stockholder. Characters: 24, 34, 59, 66, 68.

CASE E: *J. Simpson Dean*[35] contemplated whether Dean and his wife were required to pay certain income tax in view of an interest-free loan granted to them by the Nemours Corporation. The Supreme Court held no. Characters: 3, 7, 9, 13, 18, 28, 34, 61, 66.

CASE F: *Commissioner v. Duberstein*[36] concerned the case in which Duberstein received a Cadillac as a gift from Berman. Berman stated that the Cadillac was a gift in return for business favors of Duberstein, but deducted the cost as a business expense. The Court held that the car represented income to Duberstein. Characters: 4, 5, 8, 12, 15, 18, 28, 30, 31, 34, 55, 58, 60, 62, 63, 64.

These cases will be used to illustrate the basic concepts of *monothetic* and *polythetic* classes of cases.

The small selection of characters which follows illustrates the concept of *polythetic*.

CHARACTERS	*CASES*				
	A	B	C	E	F
6 Capital	1	0	1	0	0
13 Consideration	0	1	0	1	0
28 Gift	0	1	0	1	1
31 Gross	1	0	1	0	1

In this example, the defining set P of characters consists of the set 6, 13, 28, and 31. Each of the cases from the set A, B, C, E, F heads a column which determines whether or not a given character occurs in the header case. Thus, Case A contains the character 6 [Capital] denoted by the presence of the 1 in the first row of column A while it does not contain the character 13 [Consideration] indicated by 0 in the second row of column A. Case B contains the character 13 [Consideration] as indicated by the appropriate 1 in row 13 of column B, but does not contain 6 [Capital] as indicated by the 0 in row 6 of column B.

Note that the table reveals the following three properties.

34. 119 F.2d 623 (3d Cir. 1941).
35. 35 T.C. 1083 (1961).
36. 363 U.S. 278 (1960).

i) each column, that is, each case contains fifty percent (a substantial percentage) of the characters,
ii) each character is contained in a large number of cases, and
iii) no character in P is contained in all the cases.

This is an example of a class of cases which is both *polythetic* and *fully polythetic*.

A second table has been constructed to illustrate the concept of *monothetic* using the same collection of cases, but now restricting P to only the two characters 34 [Income] and 68 [Tax].

CHARACTERS	*CASES*					
	A	B	C	D	E	F
34 Income	1	1	1	1	1	1
68 Tax	1	1	1	1	1	1

Each of the cases deals with the federal income tax. More precisely, these cases A through F form a small selection from the *monothetic* class of federal income tax cases, this last class containing a case if, and only if, it has the characters 34 and 68.

These examples are intended as an explanation, not only of what is meant by a monothetic or polythetic class of cases, but also as examples of what is meant by a *legal taxonomic relationship*. The concept of a legal taxonomic relationship can be greatly refined by attaching a certain legal meaning to various standard words from biological taxonomy such as: phyletic, phenetic, cladistic, and a generic. However, this description should be adequate to indicate the general direction of taxonomic research.

TABLE 2

KEY CHARACTERS FOR INCOME TAX CASES

1	Account	38	Internal
2	Antitrust	39	Joint
3	Assignment	40	Legal
4	Awards	41	Legislative
5	Benefit	42	Limitation
6	Capital	43	Loan
7	Case	44	Moral
8	Code	45	Motive
9	Company	46	Nondeduction
10	Compensation	47	Note
11	Commission	48	Obligation
12	Compensated	49	Ordinary
13	Consideration	50	Payment
14	Corporation	51	Pension

15	Customer
16	Damages
17	Deduction
18	Deficiency
19	Donative
20	Employer
21	Exemplary
22	Exclusion
23	Exempt
24	Expenditure
25	Fraud
26	Free
27	Gain
28	Gift
29	Grace
30	Gratuity
31	Gross
32	Holding
33	Illegal
34	Income
35	Indebtedness
36	Intent
37	Interest
52	Period
53	Personal
54	Possession
55	Prizes
56	Profits
57	Receipts
58	Refund
59	Rental
60	Resign
61	Return
62	Retirement
63	Revenue
64	Salary
65	Shares
66	Stock
67	Stockholder
68	Tax
69	Taxable
70	Treasure
71	Trove
72	Trust
73	Waiver
74	Windfall

The general philosophy of numerical taxonomy generates a number of classifying techniques, many of which are well described in the very useful book by John Hartigan of Yale on clustering algorithms.[37] Among other devices, Hartigan lists profiles, distances, quick partition algorithms, k-means algorithms, partition by exact optimization, drawing trees, and single-linkage trees. Each of these have applications to the classification and organization of complex data such as case law. Researchers have developed complicated computer packages, for example at SUNY Stony Brook, for carrying out computations on data introduced in a standard format.

In spite of a high level of activity in the general area, Mackaay and Robillard appear to have made perhaps the only application to the law.[38] Their method depends upon the introduction of a *distance function* to compare and distinguish cases on the basis of a well-determined set of properties. They use the term *descriptor* in much the sense that this paper and most lawyers tend to use the term *issue*. Nevertheless, the reader is warned that these authors use the word in a more or less technical sense which itself has constituted a

37. J. HARTIGAN, CLUSTERING ALGORITHMS 1975.

38. Mackaay & Robillard, *supra* note 1, at 302.

topic of research for some investigators.[39]

Mackaay and Robillard provide a convenient outline of a "general method" of predicting judicial decisions which can be presented as follows: (1) select a line of cases dealing with the problem at hand, (2) determine a set of descriptors (issues) by means of which the fact patterns of each case can be delineated, (3) make a careful legal analysis of the cases to fix the fact pattern of each with respect to these descriptors, (4) use any ingenious scheme which comes to mind to enhance these results, and (5) apply your prediction procedure to the results of (4).

They assert that the less homogenous the line of cases, the more general the issues and the less precise the prediction, and note that Lawlor has developed useful procedures for making key selections.[40] Naturally, it is commonplace to examine the case in the usual lawyering fashion for such a determination of issues. For the purpose of linear analysis, the number of cases should be substantially larger than the number of issues mentioned. Given otherwise, Mackaay and Robillard recommend factor analysis, remarking that Kort was one of the earliest to use it in legal analysis and noting that Lawlor suggested intuitive regrouping and scaling as an alternative.[41]

Before returning to the specifics of the distance function used by Mackaay and Robillard, perhaps it may be informative to note that a variety of distance functions have been used in numerical taxonomy and cluster analysis, some of which have a statistical basis and some of which do not. Both Hartigan and Sokal discuss the choice of a distance function and related problems from a number of points of view.[42]

A case may be described by means of an n-dimensional, in this case when using Table 2, a 74-dimensional vector $(x_1, x_2, \ldots, x_{74})$ where $x_1 = 1$ if account is an issue and $x_1 = 0$ if it is not; $x_2 = 1$, if anti-trust is an issue and $x_2 = 0$ if it is not; . . . ; $x_{74} = 1$ if windfall is an issue and $x_{74} = 0$ if it is not. In these terms, two cases A and B are described by two vectors $(x_1, x_2, \ldots, x_{74})$ and $(y_1, y_2, \ldots, y_{74})$, respectively. Hamming defines the "distance" from Case A to Case B to be the number of places in which the two vectors differ and de-

39. *See, e.g.*, R. Lawlor, Applied Jurimetrics—Case Law Analysis Manual (1969) (unpublished paper); S. NAGEL, THE LEGAL PROCESS FROM A BEHAVIOURAL PERSPECTIVE, chs. 9 & 13 (1969).

40. Mackaay & Robillard, *supra* note 1, at 303-05, and elsewhere for a more nearly complete set of references to the important work of Lawlor in this area.

41. *Id.* at 303,305.

42. P. SNEATH & R. SOKAL, *supra* note 27, ch. 4; J. HARTIGAN, *supra* note 37, ch. 2.

notes the *Hamming Distance* by D(A, B).[43] Investigators find a number of advantages and disadvantages in using this distance function, however, in many respects the particular form has little effect on the clustering procedures developed around it, *i.e.*, one can use different distance functions with the same clustering algorithms.

Mackaay and Robillard developed a natural prediction procedure based on a line $A_1, \ldots, A_{64}$ of sixty-four cases concerned with capital gains taxes in Canada.[44] By means of legal analysis of these cases, they selected a collection of forty-six key issues so that each case A_i of the 64 determines a *description vector* V_{A_i} of a 46-dimensional description space E_{46}. Any new case A provides a new description vector V_A for which their program determines the *nearest* neighbour, that is, that vector V_{A_i} which is nearest to V_{A_i} with respect to the Hamming distance. The predicted outcome for A is the actual outcome for A_i.

Of course, any scheme such as this reveals certain limitations in practice and serves as a basis for a better one. Consequently, the two researchers incorporated various improvements in their method which are reported in their paper.[45] Probably the most impressive result of their investigation is a comparison of the predictions of a linear model of Lawlor, those of the method of nearest neighbours, and those of an experienced tax attorney. The results are very favorable and those decisions on which the prediction went wrong can frequently be classified as unusual.[46]

IV. CONCLUSION

This article has suggested the possibility of developing practical methods for the prediction of judicial decisions by mathematical methods. Two of the methods discussed, the method of linear models of Haar, Sawyer and Cummings and the method of nearest neighbours of Mackaay and Robillard have shown themselves useful in practical applications. In particular, the method of Haar and his collaborators has correctly predicted ninety-nine percent of the decisions in over a thousand cases, so that in the area of prediction of Zoning Amendment cases there is remarkably little hope for improvement. Such success provides both a real opportunity and a serious need for developing linear models in other special areas, not only to verify empirically that the method works in general, but also to provide additional predicting models for the legal profession.

43. Mackaay & Robillard, *supra* note 1, at 307.
44. *Id.* at 327.
45. *Id.* at 308.
46. *Id.* at 310.

While not reaching the level of precision of Haar, the method of nearest neighbours of Mackaay and Robillard has proved excellent as a predictor of capital gains cases in Canada. Continued research in the area of nearest neighbours can follow well-defined paths already pursued to great depths by investigators in the biological and medical sciences.

Consequently, there apparently exists opportunity for a large amount of research in the applications of numerical taxonomy to the law. The utility of catastrophic models remains to be established.