18

Decision Making and Evolution

E. C. ZEEMAN

INTRODUCTION

We shall compare decision making with biological evolution and discuss some of the similarities and differences between them. The aim is to shed light on social evolution, because the structure of society is sometimes changed by decision makers but at other times it seems to adapt to socioeconomic pressures in a manner more similar to the way a biological species evolves.

The discussion is based on the application of catastrophe theory (Thom 1972, Zeeman 1977) to Bayesian decision theory (Smith, Harrison, and Zeeman 1981). Bayesian theory is one of the standard planning tools used in industry and government. The decision x lies in some space X of possible decisions. The choice of decision is based on two things: first, information or beliefs about the future, and second, the utilities or preferences of the decision maker. I will explain in the next section how the information and utilities can be expressed mathematically as functions and integrated together to give a risk function R on the decision space X. The decision maker then chooses the minimum of R, in other words selects the decision x that minimizes the risk $R(x)$. This is called the *Bayes decision* (see Figure 18.1).

Meanwhile, in biological evolution the space X represents the possible mutations of a species. The function R represents unfitness, and is based on the likelihood of the future environment and the adaptability of possible mutants

THEORY AND EXPLANATION
IN ARCHAEOLOGY

ISBN 0-12-586960-6

to that environment. Darwinian natural selection then causes the species to evolve to a local minimum of R, representing minimal unfitness or, in other words, maximal fitness.

The similarity between the two processes is that they both select a minimum of R; the difference between them is that in the first case the minimum is *global*, whereas in the second case it is only *local*. The reason that the minimum is global in the first case is that the decision maker has access to the risk function defined over the whole of the decision space and can therefore select the global (or absolute) minimum. By contrast, in the second case mutation is only local, and if a species is already at a local minimum then natural selection will act against mutants to keep it there. If there is another lower minimum, the species will be denied access to it, even though it offers a better chance of survival. In this sense biological evolution appears to act blindly, because it cannot see the valley over the next hill.

It is a pertinent question to ask whether the evolution of human society can act equally blindly. For it is a familiar paradox to see society drifting toward some impending catastrophe or other, with the individuals in that society aware, yet apparently powerless to prevent it. By the word *catastrophe* in this context we mean some discontinuity in the structure of society brought about by gradually changing circumstances. At first sight it is not clear why gradually changing circumstances should produce a discontinuous effect—indeed it violates the intuition, since continuous causes normally produce continuous effects. However, there has been a considerable advance in the mathematical understanding of such phenomena during the last decade, and the method of modeling them is called *catastrophe theory*. The originator of the theory, René Thom (1972), chose the name to emphasize the unexpectedness of the discontinuities. Assuming very general hypotheses, there are theorems classifying the types of discontinuity that can occur, and if a phenomenon satisfies these hypotheses, then it can be modeled by one or another of a few standard geometric shapes. Although the proofs of the theorems are difficult for the nonspecialist, some of the geometric shapes are easy to visualize, as will be shown.

Let us return to the decision maker and ask what governs changes of decision. As time progresses, both the information and the utilities may be gradually changing, and consequently the risk function R will be gradually changing. Assuming that the decision maker abides by the rule of minimizing the risk, then this in turn will cause the decision x to vary in X. For most of the time this variation will be continuous, but at certain moments there may be abrupt switches of decision. For example, suppose a decision maker is holding the decision x_0 at the global minimum of R and observes that the risk at another local minimum of R at a distant decision point x_1 is gradually falling relative to the risk at x_0. Then as soon as $R(x_1)$ falls below $R(x_0)$ the decision maker will switch from x_0 to x_1 in order to abide by the rule of always minimizing the risk. This is illustrated in Figure 18.1 (a). Here the decision space X is represented by the horizontal axis, and the risk R by the vertical axis. The gradually changing risk function is illustrated by the sequence of seven graphs, labeled 1, 2, . . . , 7. Notice that the way we have drawn the graphs the risk is in fact steadily rising everywhere, including at x_0 and at x_1, but nevertheless $R(x_1)$ is falling *relative* to $R(x_0)$. The switch from x_0 to x_1 takes place at the time of graph number 4. In the language of catastrophe theory, the *decision switch* takes place at the *Maxwell point* 4, where the two minima are at the same level, and the *decision path* is said to obey the *Maxwell convention* (Thom 1972, Zeeman 1977).

The reader may protest that these graphs, which we have drawn to illustrate the point, are perhaps somewhat artificial; in fact the opposite is the case, because such graphs arise naturally from integrating together typical hypotheses about information and utilities, as we shall see in Example 2 and Figure 18.8.

In practice the decision maker may delay making the switch due to investment in the previous decision: For example, an industry may delay changing production due to investment in a plant, or a government may delay changing policy due to investment in credibility. This type of delay can be incorporated into the risk function, or specified in terms of thresholds that the excess risk must reach before the switch is made. However, for simplicity we shall ignore such thresholds in this chapter, because we want to contrast decision switches with the more fundamental type of delay that occurs in evolution.

Figure 18.1 (b) illustrates the analogous situation in biological evolution. Here the seven graphs represent a gradually changing unfitness function, caused, for example, by a gradually changing environment. The species starts at the local minimum at x_0 and will be held there by natural selection as long as that minimum exists. The resulting behavior is different from that of the decision maker because, for instance, by the time of graph number 5 the species will still be held at x_0, in spite of the fact that the minimum at x_1 is already lower. Therefore the switch will be delayed until graph number 6, where the minimum at x_0 coalesces with the maximum and disappears. The disappearance of the minimum causes a breakdown in the stability of the species, because now any mutation toward x_1 will be fitter, and so natural selection will automatically cause the species to evolve rapidly in that direction until it reaches x_1. Of course the evolution $x_0 \rightarrow x_1$ may in fact take many generations, but it is still "rapid" when compared with the evolutionary time scale, and it is liable to produce a discontinuity in the fossil record. Such a discontinuity is sometimes called a quantum evolution (Dodson 1972). In the language of catastrophe theory, the *evolution switch* takes place at the *bifurcation point* 6, where the maximum and minimum coalesce and the *evolution path* is said to obey the *delay convention* (Thom 1972, Zeeman 1977).

It is somewhat surprising to find that the same underlying mechanism of natural selection is responsible for two such manifestly different types of evolution as gradual adaptation and sudden switches. Indeed some present-day biologists, when they find that the fossil record consists of periods of relative

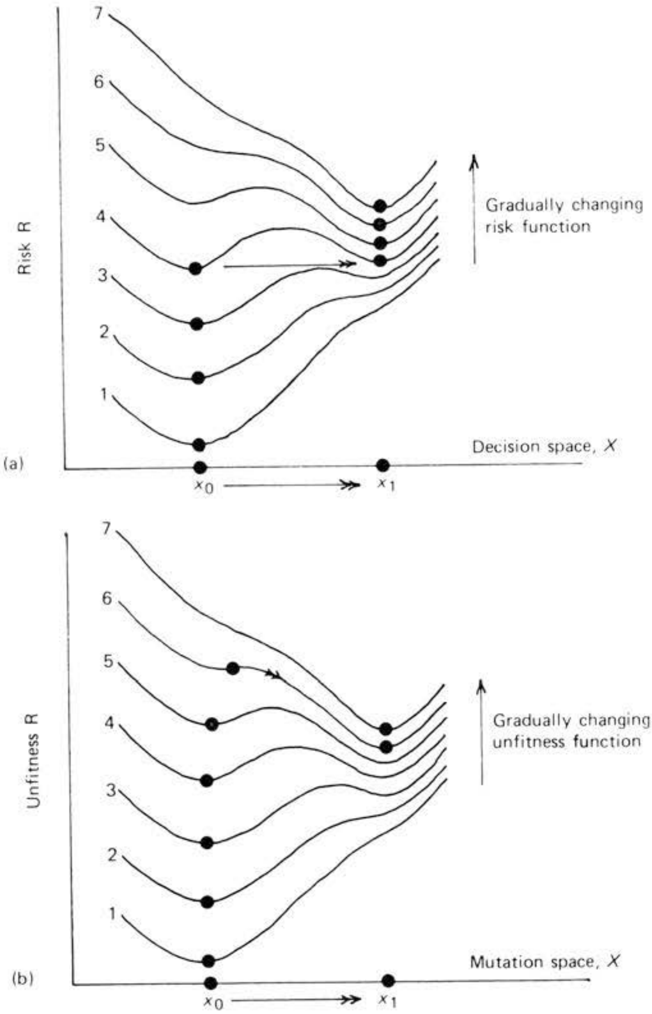

Figure 18.1(a) Graphs 1–7 show a gradually changing risk function. The resulting Bayes decisions are shown by dots. The decision switch $x_0 \rightarrow x_1$ occurs at graph number 4, where the two minima are at the same level. (b) Graphs 1–7 show a gradually changing unfitness function. The resulting evolution path is shown by dots. The evolution switch $x_0 \rightarrow x_1$ occurs at graph number 6, where the minimum at x_0 coalesces with the maximum and disappears.

constancy separated by discontinuities, mistakenly conclude that Darwin must have been wrong. On the contrary, the geometry of Figure 18.1 (b) shows that when Darwin's original concept of natural selection is applied rigorously, it predicts precisely that type of fossil record. When a fossil species terminates, indicating that the species became extinct, it does not necessarily mean that that evolutionary line died out, because it may have survived by means of an evolution switch into what appears to be a different species. Birds may be the surving dinosaurs.

SUMMARIZING:

Gradually changing risk →	Gradual adjustment of decision, most of the time Sudden switches of decision, sometimes
Gradually changing unfitness →	Gradual evolution, most of the time Sudden switches of evolution, sometimes

The comparison between the decision path and the evolution path is shown in Figure 18.2. Here the horizontal axis represents a one-dimensional parameter space C parametrizing the gradually changing risk and unfitness functions pictured in Figure 18.1. The vertical axis represents the decision space or mutation space X. The curve S represents the set of critical points of the risk and

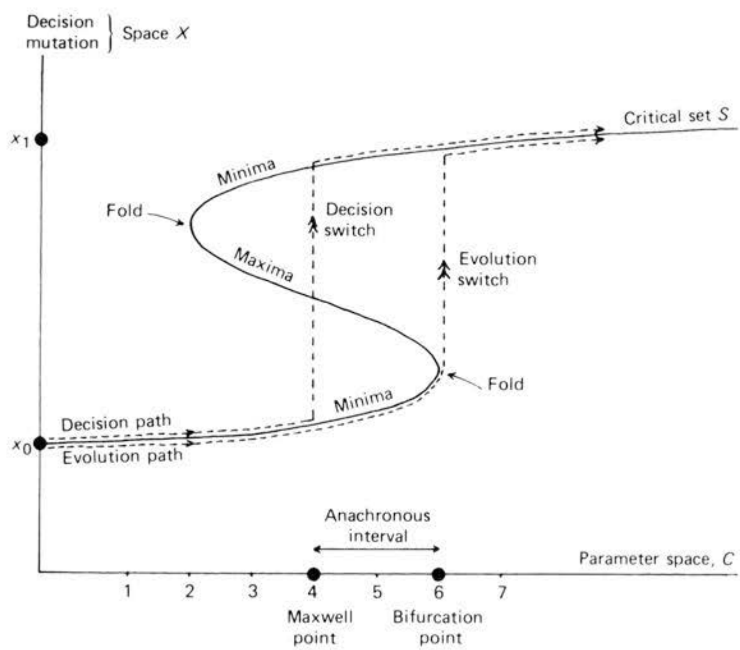

Figure 18.2. Decision and evolution paths due to a gradually increasing parameter. The decision path switches at the Maxwell point 4. The evolution path delays during the anachronous interval 46 and switches at the bifurcation point 6.

unfitness functions, the upper and lower branches representing minima, and the middle folded-over piece representing maxima. For example, over the parameter point 1 there is a single point on S corresponding to the single minimum at x_0 of graph number 1 in Figure 18.1. Meanwhile over the parameter point 3 there are three points on S, the upper and lower points representing the two minima of graph number 3 in Figure 18.1, and the middle point representing the maximum between.

The reader may wonder why I have drawn X horizontally in one picture and vertically in the other. There are good reasons: In Figure 18.1 it is natural to draw X horizontally because the risk R is a function of X. In Figure 18.2, on the other hand, we want to think of C as the *cause* and X as the *effect*, and so it is natural to draw C horizontally and X vertically. The curve S is then the graph of cause and effect, but what is unusual about this graph is that it is folded over, and it is this quality that is essentially responsible for the switches.

If the parameter is gradually increased then the decision and the evolution follow slightly different paths as shown by the dotted lines. Both start on the lower branch of S and then switch to the upper branch, but the decision switch occurs at the Maxwell point 4 while the evolution switch delays until the bifurcation point 6. We call the interval between the two switches the *anachronous interval*, because it is this interval that characterizes the difference between the two paths.

Let us examine in more detail the mechanisms that must underlie the two processes. In each case there is a *local mechanism*, and in the decision case there must also be a *global mechanism.* A local mechanism always acts by moving x in a direction that will locally reduce R. Therefore it is responsible for *stability,* because it will hold x stably at a local minimum, and if that local minimum moves, it will move x along with the local minimum. If the local minimum disappears at a bifurcation point, then the local mechanism will cause an evolution switch. Therefore a local mechanism is always responsible for both stability and evolution switches. After an evolution switch, the local mechanism will hold x stably in the new minimum, and even if the parameter is moved back across the bifurcation point where the switch occurred, it will still hold x stably in the new minimum. For instance in Figure 18.2 the parameter has to be moved right back to point 2 before the reverse evolution switch will occur at the other fold point. The difference between the two bifurcation points 2 and 6, where the opposite evolution switches take place, is called *hysteresis,* and this is a characteristic feature of local mechanisms. In the evolution case the local mechanism is natural selection, and this is the only mechanism.

By contrast, in the decision case there is both a local mechanism and a global mechanism. The local mechanism is the procedure whereby a decision maker continually explores local variations of his current decision and continually adjusts it to keep the risk at the local minimum. For example, most financial policies operate on this principle, by making incremental adjustments on the previous budget. Meanwhile the global mechanism is a much more complicated affair because it involves three ingredients: First, it involves globally exploring the risk function in order to find the global minimum. Second, it involves overriding the local mechanism at the Maxwell point in order to break the stability of the old decision and make the switch to the new decision. Third, it involves reinforcing the stability of the new decision in order to prevent a reverse switch back again and to allow time for the temporarily overridden local mechanism to reestablish itself as the natural stabilizing force at the new decision. If there is a global mechanism, the decision switch always preempts the evolution switch that would have occurred had there been no global mechanism, as can be seen from Figure 18.2. In contrast to the hysteresis at an evolution switch, there is no hysteresis at a decision switch, because if the parameter is moved back again then the reverse switch occurs at the same Maxwell point. Therefore, after a decision switch the new decision is vulnerable to reversal, especially if there is any stochastic noise in the parameter, which explains why the third ingredient of the global mechanism is necessary. The capacity to perform these three ingredients demands intelligence, and this is of course the main difference between decision making and biological evolution.

Let us now apply these ideas to *social* evolution. It is tempting to speculate (Renfrew 1978, 1979) that certain discontinuities in the archaeological record may have been caused by switches in social evolution, just as certain discontinuities in the fossil record may have been caused by switches in biological evolution. In this case, X would be a multidimensional space representing possible structures of society, and R an unfitness function that was gradually changing due to variations in population, environment, resources, technology, culture, etc. The local mechanism would be socioeconomic pressure, and this would continually adapt the structure of society to the changing conditions, just as natural selection continually adapts a biological species. An evolution switch in the structure of society would be triggered by the breakdown of the stability of an existing structure, and then the very same local mechanism of socioeconomic pressure would cause a rapid evolution to a different structure.

As yet we have not introduced intelligence, nor a global mechanism, which is the essential difference between social and biological evolution. In the social case individuals can foresee an approaching catastrophic switch, especially during the anachronous interval, during which the old structure of society is still stable and the new structure is both potentially stable and fitter. Consequently society tends to appoint decision makers who can preempt the evolution switch by means of a decision switch. Instead of having to suffer the instability that must necessarily follow the breakdown of the old structure, they can utilize the remaining stability of the old structure during the anachronous interval to usher in the new structure before the stability of the old structure has broken down.

The three ingredients that the decision makers need to establish a global mechanism are as follows. First they must collect information about alternative structures of society, decide utilities, and choose the structure that maximizes fitness (or minimizes risk). Second, they must override the local socioeconomic pressure in order to make the switch. Third, they must reinforce the stability of the new structure in order to allow time for socioeconomic pressure to reestablish itself there as the natural stabilizing force. In other words they must introduce the three familiar branches of government—the executive, the legislature, and the judiciary. In primitive societies it was often found efficient to embody all three branches in a single decision maker, such as a king or chieftain. However, as society became more complex more decision makers were needed, leading to democratization and the separation of the three branches of government.

In some societies the rulers have identified themselves too strongly with the local mechanism maintaining an existing structure and have failed to perceive the increasing unfitness of that structure, particularly if it involved too great an imbalance in the distribution of resources. As a result, society has tended to throw up other decision makers in the form of revolutionaries. The revolution is then the switch, and the three ingredients of the global mechanism that the revolutionaries need are to plan, execute, and consolidate the revolution. Since it usually takes some time for the socioeconomic pressure to reestablish itself again as the stabilizing force after a revolution, successful revolutionaries are aware of their vulnerability to the reverse switchback, and so as part of the third ingredient of consolidation they tend to be particularly severe toward counter-revolutionaries.

Some historians suggest that on the whole man's social structures are surprisingly ephemeral and short-lived, but here they may be making the same mistake as the biologists. For we should expect gradual changes of population, resources, and culture to produce periods of continuity separated by discontinuities in the historical record, analogous to those in the fossil and archaeological records. And just an an evolutionary line may survive through many switches of species, so a society may survive through many switches of structure.

So must for generalities: I shall now give the mathematics and then a number of examples. I begin by describing the general Bayesian model for decision theory. When parameters are introduced, the Bayesian model falls naturally into the mathematical domain of catastrophe theory, where there are theorems classifying the geometric shapes of decision sets. The main purpose of the examples is to demonstrate the potential of this method of modeling. A secondary purpose is to treat each example as a speculative model in its own right. However, to do each example justice would require considerably more detail in defining terms, presenting supporting evidence, specifying methods of measurement, and collecting and fitting data, than would be possible within the scope of this chapter.

Examples 1 and 2 are meant to illustrate the mathematics and show how quite simple assumptions can lead into the subleties of the cusp catastrophe. They concern predictions in the case of a one-dimensional spectrum of choice. For example, an investor might be trying to predict the price of a particular stock on the stock market in the face of ambiguous information. A government might be trying to predict its budget or a primitive society the best allocation of its resources. In each case the information is presented to the decision maker in the form of a probability distribution with a single peak, but if the distribution is skew, then the information will be ambiguous in the sense that the decision maker will not know whether to follow the mean or the mode. The ambiguity is resolved by a suitable choice of utilities, and the model gives rise to a cusp catastrophe that reveals the divisions and switches among the decision makers.

Example 3 is a more abstract model concerning the evolution of bird's beaks. Changes in the food supply cause evolution switches between specialist beaks filling ecological niches and general-purpose beaks that are suitable for a variety of foods.

Example 4 is an analogous model concerning the evolution of roles within a society, with changes in the resources and technology causing the evolution of specialist roles filling sociological niches and general purpose roles that are suitable for a variety of activities. The paradox here is that individuals can decide to switch to fitter roles, but the roles themselves tend to evolve blindly like the bird's beaks. Thus social evolution can simultaneously reflect aspects of both decision making and biological evolution.

Example 5 is an analogous model concerning the evolution of the structure of society. The main difference here is the existence of a global mechanism that eliminates anachronous structures, whereas in the previous example the harmless anachronous roles were allowed to survive.

Example 6 is a more specific version of the last example, concerning the policies put forward by political parties at elections, and discussing the splits that can occur between specialist policies and more general platforms aimed at broader constituencies.

Example 7 describes the confusion between two decision makers with different utilities and hence different risk functions, as exemplified by the misunderstandings that can arise between doctor and patient. In particular, the model sheds light on the delays that sometimes occur before a patient will admit he or she is ill or admit that he or she is better again. Further examples can be found in Harrison and Smith (1979) and Zeeman (1980).

BAYESIAN THEORY

In the next two sections I describe the general mathematical model, and in Examples 1 and 2 I give simple illustrations of the mathematics. The reader may find it helpful to read the examples alongside the general model. The model will apply to both decision making and evolution, but for simplicity I phrase it mainly in terms of decision theory (Smith, Harrison, and Zeeman 1981).

Let $\mathbb{R}$ denote the real numbers (the heavy type distinguishes $\mathbb{R}$ from the risk function R), and let $\mathbb{R}^n$ *denote* n-dimensional space. The data in a Bayesian model consist of four things: $X =$ decision space; $Y =$ future space; $P =$ probability distribution; and $L =$ loss function. From this data we shall construct the following: $R =$ risk function.

The *decision space* X is the set of possible decisions x. In some examples we shall have a single spectrum of decisions, $X = \mathbb{R}$. In other examples the decisions may depend on n variables, so that X will be an open subset of $\mathbb{R}^n$. Indeed in some applications it is necessary to have n large in order to represent all the multiple relevant interacting factors involved, but then the mathematical theorems will allow us to leave most of these variables implicit, as is explained below.

The *future space* Y is a set of future possibilities y. These possibilities may refer to some specific time in the future, or they may represent future developments over a specified period. In some examples there will be a single spectrum of possibilities $Y = \mathbb{R}$, but in other examples Y may be multidimensional, or more complicated. The only mathematical requirement on Y is that it should have a measure, so that we can integrate over it to cover all possibilities. The particular measure on Y is unimportant because the risk function turns out to be independent of the measure. This is useful because different decision makers can then use different measures, but they will arrive at the same decision, provided they have the same information and utilities. If the decision is merely a prediction of the future then we shall have $X = Y$. On the other hand, Y may be much more complicated than X, because the future may hold many more possibilities than there are options open to the decision maker. Indeed the very purpose of Bayesian decision theory is to enable the decision maker to allow for the complexity of the future.

The *information* about the future, or the decision maker's belief about the future, is contained in the probability distribution $P: X \times Y \rightarrow \mathbb{R}$. Here $P(x, y)$ is the probability that if the decision x is made then the future y will occur. More precisely, if dy is a measure-element of Y at y, then $P(x, y)dy$ is the probability that the future will lie in dy. Therefore, for each x in X,

$$\int_Y P(x, y)\,dy = 1.$$

If the decision has no effect upon the future, then P will be independent of x, $P(x, y) = P(y)$.

The *utilities* or preferences of the decision maker are contained in the *loss function* $L: X \times Y \rightarrow \mathbb{R}$. Here $L(x, y)$ is defined to be the loss that the decision maker will incur if decision x is made and then the future y subsequently happens.

Having been given the data we can now define the *risk function* $R: X \rightarrow \mathbb{R}$, as follows:

$$R(x) = \int_Y L(x,y)P(x,y)\,dy.$$

In other words the risk function is the expected loss: The decision maker weights each future probability-element $P(x,y)dy$ by the appropriate loss $L(x,y)$ and then integrates over Y to cover all possibilities in order to obtain the expected loss for that decision. We assume that the resulting function R depends smoothly on x (see the Appendix).

Call x a *critical point* of R if the gradient ∇R of R vanishes at x. Generically the critical points of R will be minima, maxima, and saddles (if $n > 1$). However nongeneric critical points may occur generically in a parametrized family of R's, as for example the point where the maximum and minimum have coalesced on graph number 6 in Figure 18.1 (b). Such points are fundamental in catastrophe theory because they are responsible for the evolution switches. Let $S =$ the set of critical points, given by $\nabla R = 0$; $E =$ the subset of minima; and $D =$ the subset of absolute minima. Then $D \subset E \subset S \subset X$, and generically D will contain exactly one point that is the *Bayes decision* minimizing the risk. However, it may occur generically in a parametrized family of R's that some D's will contain more than one point, as for example in graph number 4 in Figure 18.1 (a) where both minima are the same level. Such points are also fundamental in catastrophe theory because they are responsible for the decision switches.

CATASTROPHE THEORY

We now formally introduce the parameters. Let C be a *parameter space* governing P and L. For example if P and L depend upon k continuous parameters, then C will be an open subset of $\mathbb{R}^k$. Usually k is small because we are interested in what happens to the decision when we change only a few parameters.

For each point c in C, we are given a probability distribution P_c and a loss function L_c, from which we can calculate the risk function R_c, and hence deduce the critical points S_c, minima E_c, and absolute minima D_c. Now let c vary in C and define the

critical set $S = \{(c, x); S_c$ contains $x\} =$ all critical points,
evolution set $E = \{(c, x); E_c$ contains $x\} =$ all minima,
decision set $D = \{(c, x); D_c$ contains $x\} =$ all absolute minima.

Then

$$D \subset E \subset S \subset C \times X.$$

If L is generic for P (see the Appendix for a precise definition) then S turns out to be a smooth k-dimensional submanifold of $C \times X$, the same dimension as C, and independent of the dimension of X. Furthermore E and D are submanifolds of S of the same dimension. The significance of E and D are that evolution paths lie in E and decision paths lie in D, and the respective switches occur whenever the paths cross the boundaries of these submanifolds.

For example in Figure 18.2, $C \times X$ is two-dimensional, and the critical set S is the one-dimensional S-shaped curve. The evolution set E consists of the upper and lower branches of S, and is obtained from S by removing the folded-over middle piece. The decision set D is obtained by further removing the two short pieces of E not in the decision path. In other words D is the intersection of E with the dotted decision path. In Figure 18.10, $C \times X$ is three-dimensional and S is the two-dimentional folded surface. E is obtained from S by cutting along the fold curve and removing the folded-over middle piece. D is obtained by cutting along the dotted line and further removing the two pieces in between the dotted line and the fold curve.

Let χ denote the map projecting S onto C, which is called the *catastrophe map* because it is mathematically important. Let ∂E, ∂D denote the boundaries of E, D and define their projections in C to be the

$$\textit{bifurcation set } B = \chi(\partial E),$$
$$\textit{Maxwell set } M = \chi(\partial D).$$

The significance of B and M is that evolution switches occur at bifurcation points in B, and decision switches occur at Maxwell points in M. For example, in Figure 18.2 ∂E consists of the two fold points of S, and the bifurcation set B consists of the two points 2 and 6 beneath them on the X axis. Meanwhile ∂D consists of the two ends of the decision switch, and the Maxwell set M is the single point 4 beneath them. In Figure 18.10 ∂E is the fold curve, and the bifurcation set B is the cusp beneath it; ∂D is the dotted line on S, and the Maxwell set M is the dotted line beneath it inside the cusp.

One of our main objectives is to understand what types of decision and evolution paths are possible when the parameters are varied, in order to fully comprehend the relationship between gradual changes and sudden switches. This is where catastrophe theory comes to our assistance because its theorems classify the possible shapes of S. More precisely they classify the singularities of the catastrophe map $\chi: S \to C$.

What is a singularity? A point of S is defined to be a *singularity* of χ if it has a vertical tangent.[1] For example in Figure 18.2 the singularities are the two fold points. In Figure 18.10 the singularities are the points along the fold curve, and the point vertically above the cusp point. In general, the set of singularities contains ∂E and other points (but the other points, such as those bounding the set of maxima, are less important because they do not represent evolution switches if $n > 2$).

Most points of S are nonsingular, and a characteristic property of a nonsingular point is that it has a neighborhood in S that is mapped by χ into C in

[1]Here *vertical* means parallel to X. This definition is equivalent to the usual mathematical definition that the derivative of χ drops rank.

a one-to-one manner. Consequently local variations of the parameters at a nonsingular point can only cause local variations of evolution, and so the evolutionary state is stable. By contrast, a singularity has no such neighborhood, and so a small variation of the parameters is liable to cause an evolution switch. If there were no singularities, then there could be no evolution switches. Nor could there be any decision switches if S were connected, because then χ would have to map S onto C in a one-to-one manner. Therefore it is the singularities of χ that are ultimately responsible for both types of switch, and that is why we need to classify them.

If C is one-dimensional, that is to say there is only one parameter, then there is only one type of singularity, namely the *fold,* which is illustrated in Figure 18.2. If C is two-dimensional, that is to say there are two parameters, then a new type of singularity appears, namely the *cusp* illustrated in Figure 18.10. If C is three-dimensional, then three more types appear, if C is four-dimensional two more types appear, and so on.

Thom calls the structure of S surrounding one of these singularities an *elementary catastrophe*, and descriptions of the elementary catastrophes together with a proof of the classification theorems can be found in Thom (1972) and Zeeman (1977). In the Appendix to this chapter we give a precise definition of genericity of L sufficient for the theorems to hold. A fundamental property of elementary catastrophes is their stability under perturbations of P and L.

Of the elementary catastrophes the *fold* is important because it models the evolution switch. The *cusp* is important because, as we shall see later in Example 2, it models conflicting decisions. When C is four-dimensional, one of the elementary catastrophes is called the *butterfly,* and this is important because it can be used to model the emergence of compromise decision (Isnard and Zeeman 1975). When C is five-dimensional there is another catastrophe called E_6 that is being increasingly used to model various forms of psychological decision making and psychotherapy (Zeeman 1977, Callahan 1981). When C is eight-dimensional there is an important catastrophe called the *double cusp,* which already has several applications in physics and can also be used to model the interference between two conflicts (Zeeman 1977).

One of the advantages of concentrating on S rather than X is that it enables us to reduce the number of variables and make the model quantitative. For example suppose $N = 1000$ and $k = 2$, in other words there are 1000 relevant interacting factors involved in the decision, and we are interested in the effect of two particular parameters. Then $C \times X$ will be 1002-dimensional, but the theorems reassure us that S is nonetheless a smooth two-dimensional surface lying inside this 1002–dimensional space. Therefore it is possible to represent S as an ordinary two-dimensional surface in three dimensions lying over the horizontal plane C, with the vertical axis being a suitably chosen X variable. In other words, to describe S quantitatively it suffices to measure only one of the 1000 X variables explicitly and leave the other 999 implicit. Moreover, the theorems tell us that the most complicated shape that S can have locally is the cusp catastrophe shown in Figure 18.10. Globally there may be several cusps and folds, as for instance in Figure 18.11.

A *local mechanism* can be represented mathematically by a differential equation on X parametrized by C that reduces R, in other words, such that $\dot{R} < 0$ whenever $\nabla R \neq 0$, and $\dot{R} = 0$ whenever $\nabla R = 0$ (where the dot denotes the rate of change). For example, the gradient differential equation $\dot{x} = -\nabla R$ is of this type. Such a differential equation will cause x to seek E, to stay on E if the parameters are changed, and to switch to other parts of E if ∂E is crossed. Thus the differential equation provides both the stability and the evolution switches of a local mechanism.

Note that the differential equation itself is defined on the 1002-dimensional space $C \times X$, and must therefore usually remain implicit, but the resulting evolution paths can be represented explicitly on E in the three-dimensional picture. Similarly, if there is an implicit global mechanism in addition to the implicit local mechanism, the resulting decision paths can be represented explicitly on D in the three-dimensional picture.

In summary, in effect the general mathematical model has three levels of complexity. First, the possibilities Y of the future may be very complicated. Second, the possible decisions X are usually simpler than Y, but may still be highly multidimensional. Third, when we get down to the decision set D, this can sometimes be measured because it is low-dimensional, but may be subtle because of the switches. However, the switches are determined by the singularities, and they are classified by the theorems.

Example 1 : Predicting the Price of Stock

For the first example I choose something that is easy to understand and easy to calculate. Consider the problem of whether to buy or sell a particular stock on the stock market. Let x denote today's prediction of tomorrow's price, and let y denote tomorrow's price. In other words, the decision x is a prediction of the future y. Since x, y are both numbers, we have $X = Y = \mathbb{R}$.

The reader may well ask what has this to do with archaeology? In fact the same model can be applied to any situation where a decision maker is trying to predict something that can be measured. For instance, an archaeologist might be trying to predict next year's budget, or a government might be trying to predict the rate of inflation or the future food supply. And of course these are not only current problems, but must also have preoccupied the rulers of primitive tribes and ancient civilizations, who must have been faced from time to time with famines, or increasing populations and dwindling resources, and decisions about whether to hunt or to cultivate or to emigrate.

So let us return to the stock market in order to fix our minds. Suppose that the information about the future is given by the probability distribution P shown in Figure 18.3. Being only a small investor we assume that our decision x has no effect on the future y, and so P is only a function of y and independent of x:

$$P(x, y) = P(y) = \text{probability that tomorrow's price is } y.$$

We have deliberately chosen a skew distribution so that the mode m is different from the mean μ. This makes the information ambiguous, because it is not clear whether we ought to follow the mode or the mean. The mode is the most likely price tomorrow, and the mean is the expected price tomorrow. If today's, price happens to lie between them, then following the mode would suggest selling stock because the most likely price tomorrow will be lower than today's, whereas following the mean would suggest buying because the expected price tomorrow will be higher. Statistics emphasizes the importance of the mean, but stockbrokers, who are generally richer than statisticians, tend to follow the mode, so whom should we follow? The ambiguity and indecision are resolved by choosing a loss function L.

The simplest loss function is the parabola $L(x, y) = (x - y)^2$ (see Figure 18.4). If the prediction is correct, $x = y$, then there is no loss. If not, then $x - y$ is the error of prediction, and the parabola can be regarded as a translation into mathematics of the qualitative statement "*The greater the error the greater the loss.*" Using this loss function we can now calculate the risk function, which is shown in Figure 18.5.

Lemma. The risk function is a parabola with minimum at the mean μ. Hence the Bayes decision is the mean.

Proof: The risk $R(x) = \int L(x, y)P(y)dy = \int (x - y)^2 P(y)dy$.

The variance of P, $V = \int (\mu - y)^2 P(y)dy$.

$$\begin{aligned} \text{Therefore } R(x) - V &= (x^2 - \mu^2)\int P(y)dy - 2(x - \mu)\int yP(y)dy \\ &= (x^2 - \mu^2) - 2(x - \mu)\mu \\ &= (x - \mu)^2. \end{aligned}$$

Therefore $R(x) = V + (x - \mu)^2$, as required.

It is a confusing coincidence that a parabolic loss function should give rise to a parabolic risk function, because generally the two functions are quite different from one another as in our next example. The parabolic loss function can be criticized on the grounds that it is unreasonable for the loss to tend to infinity as the error increases. More commonly the decision maker has an upper bound to his losses because he cannot lose more than he possesses. Therefore it is more reasonable to assume a fixed loss for sufficiently large error, and so I have incorporated this assumption into the next example.

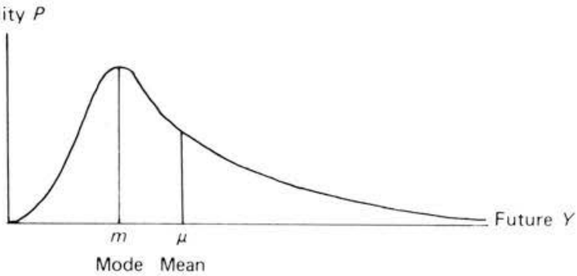

Figure 18.3. Skew probability distribution.

Example 2 : Ambiguity and Caution

Example 1 is modified by replacing the parabolic loss function of Figure 18.4 with the two-step loss function shown in Figure 18.6. As before L depends on the error but this time it is given by the formula

$$L = \begin{cases} 0, & 0 \leq |x - y| \leq a \\ 1 - \alpha, & a < |x - y| \leq b \\ 1, & |x - y| > b \end{cases}$$

where a, b, α are constants, a small, b large, and $0 \leq \alpha \leq 1$. Although L is discontinuous at $\pm a$, $\pm b$ it is nevertheless generic for the probability distribution P shown in Figure 18.3, according to the definition in the Appendix.

The two-step loss function can be regarded as a translation into mathematics of the qualitative statement "*Spot on you win; way out you lose.*" For suppose L^* is the loss function in which the decision maker makes a fixed profit p (in other words a negative loss $-p$) if the error is sufficiently small $|x - y| \leq a$, and makes a fixed loss q if it is sufficiently large $|x - y| > b$, and otherwise breaks even. Then we can normalize L^* by defining

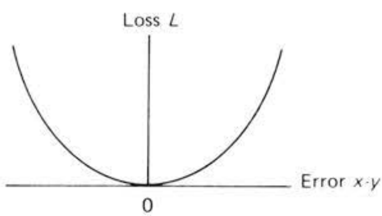

Figure 18.4. Parabolic loss function.

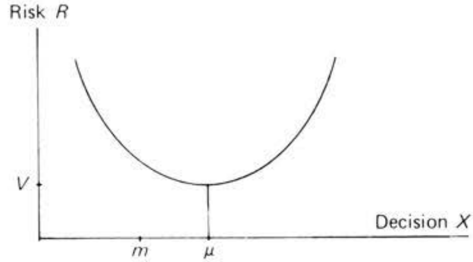

Figure 18.5. Parabolic risk function.

$$L = \frac{p + L^*}{p + q}, \qquad \alpha = \frac{q}{p + q},$$

and this gives the two-step loss function L above, shown in Figure 18.6. Such a normalization is admissible because affine changes of L (that is adding and multiplying by constants) induce the same affine changes in R, and so do not alter the critical points of R nor the Bayes decision.

If $\alpha = 0$ the decision maker is only concerned with his profits, and therefore will behave like a *speculator*, being primarily interested in increasing his capital by speculating with those stocks giving the greatest return. On the other hand, if $\alpha = 1$ the decision maker is only concerned with his losses, and therefore will behave like an *investor* who is primarily interested in conserving his capital by investing in securities. Therefore we can interpret the parameter α as a measure of the *caution* of the decision maker, varying from speculator to investor as the caution increases.

The risk function R can now be computed as follows. Regard L as the

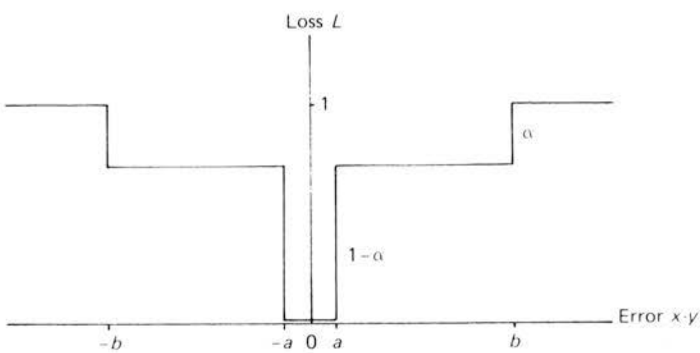

Figure 18.6. Two-step loss function.

constant function, with value 1, minus two rectangles, the first of width $2a$ and height $1 - \alpha$, and the second of width $2b$ and height α. Therefore

$$R(x) = \int L(x, y)P(y)dy = 1 - (1 - \alpha)A(x) - \alpha B(x),$$

where

$$A(x) = \int_{|x-y| \leq a} P(y)dy = \int_{x-a}^{x+a} P(y)dy$$

$$B(x) = \int_{|x-y| \leq b} P(y)dy = \int_{x-b}^{x+b} P(y)dy.$$

The critical points of A are given by

$$\frac{dA}{dx} = P(x + a) - P(x - a) = 0.$$

Therefore A has a unique maximum at m' say, where m' is the midpoint of the unique horizontal chord of P of length $2a$. Since a is small, m' is near the mode m of P. Similarly B has a unique maximum at μ' say, where μ' is the midpoint of the chord of length $2b$. If b is of the order of about twice the standard deviation of P then it can be seen from Figure 18.3 that μ' is near the mean μ of P, as shown in Figure 18.7. Therefore if $\alpha = 0$ then R has a unique minimum at m', and if $\alpha = 1$ then R has a unique minimum at μ'. If $0 < \alpha < 1$ then R is a linear combination of A and B, and so it is either unimodal, with a unique minimum in between m' and μ', or else bimodal with two minima, one near the mode and the other near the mean, as shown in Figure 18.8. It is surprising at first sight that a bimodal R can arise from a unimodal P and unimodal L, but the way we have computed R explains how this phenomenon arises, and shows it to be stable under small perturbations of P and L. As the parameter α varies from 0 to 1 the decreasing family of loss functions give rise to a smoothly decreasing family of smooth risk functions, as shown in Figure 18.8 (compare Figure 18.1).

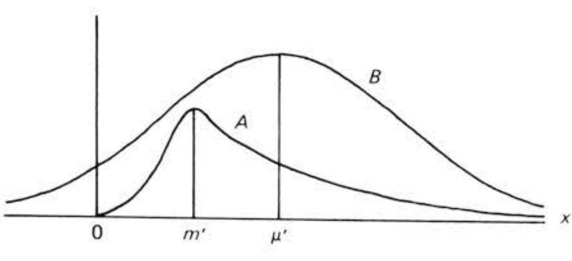

Figure 18.7. The functions A and B.

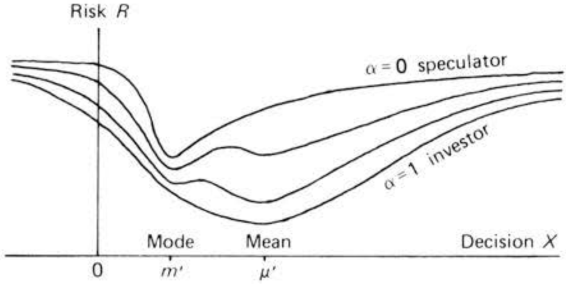

Figure 18.8. The family of risk functions.

Therefore the speculators ($\alpha = 0$) will follow the mode and the investors ($\alpha = 1$) will follow the mean. If a decision maker is gradually becoming more cautious (α increasing), then he will suddenly switch from mode to mean at the Maxwell point when the two minima are level. Conversely, if he is gradually becoming more adventurous (α decreasing) he will make the reverse switch at the same point. The bimodality of R resolves the original dilemma posed by the ambiguity of the information P as to whether to follow the mode or the mean, and shows how it depends upon the choice of loss function.

We now introduce a second parameter β that measures the *ambiguity* of the information, or in other words the skewness of P. Define

$$\beta = \mu - m = \text{mean minus mode.}$$

A symmetrical distribution like the normal distribution would have $\beta = 0$. If $\beta > 0$, then P is skewed to the right as in Figure 18.3, and if $\beta < 0$, then P is skewed to the left.

Suppose we are given a one-parameter family of distributions all skewed to the right as in Figure 18.3 and parametrized by their skewness β. Then our parameter space C will now be two-dimensional with coordinates α measuring the caution of the decision maker and β measuring the ambiguity of information. In order to find the critical set S we must investigate the modality of R_c for each parameter point $c = (\alpha, \beta)$. Given β, let A_β, B_β denote the corresponding functions of Figure 18.7, let m_β, μ_β denote their maxima, and let S_β denote the resulting section of S over the interval $0 \leq \alpha \leq 1$.

If β is small then m_β, μ_β will be too close together for there to be any points of inflexion of A_β, B_β between them, and this condition is sufficient to ensure that $(1 - \alpha)A_\beta + \alpha B_\beta$ has a unique maximum, for all α, $0 \leq \alpha \leq 1$ (Smith 1977). Therefore R_c has a unique minimum for all α, and so S_β is single-valued over α, as in Figure 18.9 (a).

If β is large, then R_c will be bimodal for some value of α, $0 < \alpha < 1$, as in Figure 18.8, and therefore S_β will be folded over as in Figure 18.9 (b) (compare Figure 18.2). Since these are the only two types of section, there must exist a critical value β_0 such that if $\beta < \beta_0$ then S_β is as in the case (a) and if $\beta > \beta_0$

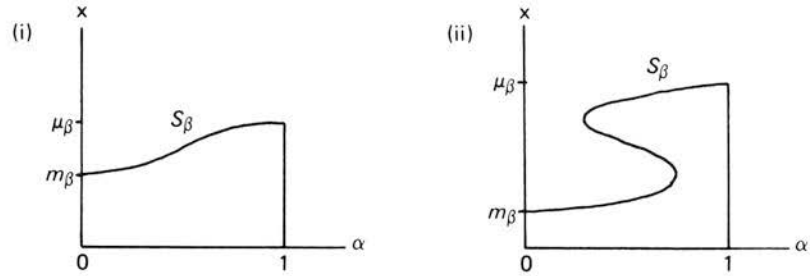

Figure 18.9. Sections of S_β for (a) small β and (b) large β.

then S_β is as in case (b). Then it is a theorem (Zeeman 1977) that the sections can be assembled into a surface S, as shown in Figure 18.10. This is called a *cusp catastrophe* because its bifurcation set is a cusp, and it is the unique elementary catastrophe for a two-dimensional parameter space. Furthermore it is stable, and so its qualitative properties are preserved under perturbations of P and L. Figure 18.10 provides a synthesis of these qualitative properties, as follows.

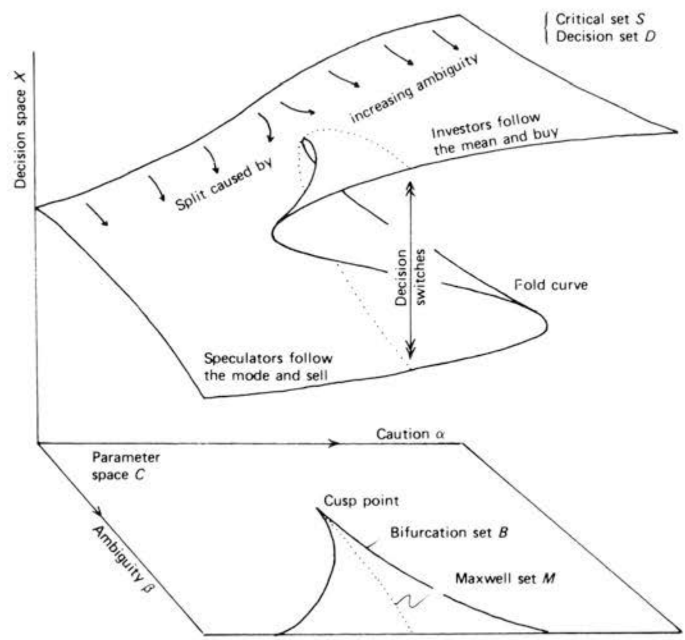

Figure 18.10. The predicted price of stock is a cusp catastrophe with the caution of the decision maker as normal factor and the ambiguity of information as splitting factor. The continuous spectrum of decision makers is split into buyers and sellers by increasing ambiguity of information. Individuals switch decision if their level of caution crosses the Maxwell set.

The decision set D is bounded by the dotted line on S vertically above the Maxwell set M, which is the dotted line inside the cusp in C. Since the decision x is correlated with the caution α we call α a *normal factor*. Since increasing ambiguity β has the effect of splitting D apart we call β a *splitting factor*. The β coordinate of the cusp point is the critical value β_0 separating the two types of section S_β shown in Figure 18.9. Therefore if $\beta < \beta_0$ the decision makers of varying levels of caution α will be spread continously along the section S_β, and so their predictions x of tomorrow's price of the stock will form a continuous spectrum. If, however, the ambiguity β is increased beyond the threshold β_0 then they will be split into two groups (as indicated by the arrows on S in Figure 18.10), with the speculators following the mode and the investors following the mean. Therefore if today's price of the stock happens to lie between the mode and mean, the speculators will begin to sell and the investors will begin to buy (which is very convenient to both parties).

If the decision maker's path in the parameter space happens to cross the Maxwell set, for instance if he changes his level of caution while believing the information to be ambiguous, then this will cause him to suddenly switch his decision from buying to selling, or vice versa.

When $\beta > \beta_0$ what is surprising is that *all* the decision makers are split, including those of moderate caution. Although some people are predicting that tomorrow's price will be higher than today's, and others are predicting that it will be lower, no one is actually predicting that it will be the same, or anywhere near the same. The switch from one group to the other jumps right over this possibility. This phenomenon is called *inaccessibility*, and is a characteristic feature of the cusp catastrophe (Zeeman 1977). It is particularly useful in modeling polarized situations, where a population may initially exhibit a continuous spectrum opinion over some issue, but if that issue increases in urgency beyond some threshold then the population may find itself split into taking sides, with a sharp division of opinion between the two sides, the intermediate ground being inaccessible (Isnard and Zeeman 1975).

If the information is skewed the other way, $\beta < 0$, implying $\mu < m$, then another symmetrically placed cusp appears, with the orientations reversed as in Figure 18.11. As the market oscillates back and forth, first skewed one way and then the other, the speculators and investors interchange their roles of buying and selling to each other. Michael Thompson (1981) has used this geometry to analyze structures of society.

Example 3 : The Evolution of Bird's Beaks

We apply a generalization of Example 2 to the evolution of bird's beaks. Consider how the shape of a beak is affected by the variety of food available and the suitability of that beak for the various types of food. Let X describe the possible shapes of beak. For example, if we use n variables to measure the

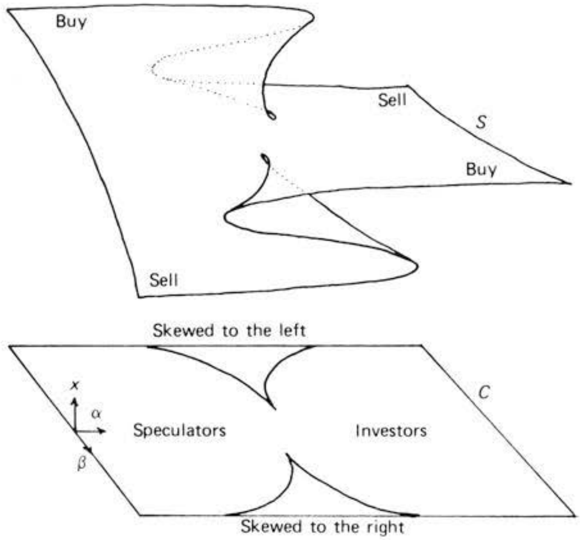

Figure 18.11. Skewing either way gives two cusps.

shape of a beak, and represent these measurements by a point x in $\mathbb{R}^n$, then X will be a subset of $\mathbb{R}^n$. Let Y denote the set of available foods, and let P be a probability distribution on Y representing the probable availability of each food in the given environment. Let the function $L(x, y)$ represent the inefficiency of beak x for food y. For instance we might invent some scale running from perfectly efficient at $L = 0$ to useless at $L = 1$. Then $R(x) = \int L(x, y)P(y)dy$ will measure the unfitness of beak x for that environment. Natural selection will be the local mechanism that evolves the beak to a local minimum of R.

If there is a dominant food supply y_0 then the integral will be dominated by $L(x, y_0)P(y_0)$, and so R will be minimized by evolving a specialist beak x_0 with the minimum inefficiency $L(x_0,y_0)$ (or maximum efficiency) for eating food y_0. In other words, the beak x_0 is specialized to fill the ecological niche y_0. However, the very success of a species in filling an ecological niche will tend to increase the number of individuals competing for that niche, and hence reduce the food available to each individual. This in turn may produce a selection in favor of some form of population control, for example, the development of a territorial instinct, which will maintain $P(y_0)$ at a sufficiently high individual level. Otherwise, the expanding population may have an effect similar to the skewing of P, by increasing the relative importance of other food supplies. Consequently, a bimodality may appear in R, as in Example 2, between the specialist beak x_0 filling the niche y_0 (corresponding to the speculator) and a general purpose beak x_1 adapted to eating a wide variety of foods (corresponding to the investor). Therefore the gradually increasing skewness of the food supply caused by the success of a species may in turn cause a bifurcation of the species into the gradual divergent evolution of two different types of beak, one a specialist and the other a generalist.

Figure 18.12 shows a cusp catastrophe that synthesizes the preceding discussion. We have assumed a two-dimensional parameter space C, with the normal factor α representing the scarcity of food, and the splitting factor β representing the size of population. Unlike Figure 18.10, the vertical axis is not X because X is multidimensional. However, we know that the critical set S is a two-dimensional surface, and if it contains a cusp then any measurement $\bar{x}$ of the beak that distinguishes between specialist and generalist will suffice for a vertical coordinate if we want to plot that cusp catastrophe as a surface in three dimensions, as in Figure 18.12.

We now consider the evolution switches. Suppose a specialist beak x_0 is filling the niche y_0, which is disappearing due to gradually increasing scarcity, caused perhaps by increasing competition from other species. When the parameter crosses the right side of the cusp then the local minimum of R at x_0 will disappear [as in Figure 18.1(b)] and then natural selection will cause an evolution switch to the general-purpose beak. Conversely a gradual growth in a dominant food supply may cause the parameter to cross the left side of the cusp and induce the reverse switch from the general-purpose beak to a specialist beak. Normally abundance encourages specialists, and scarcity encourages generalists. For instance, the abundance of plant life has encouraged many striking examples of specialists in the insect world, adapted to one type of food only.

Before we leave this example, notice the difference between Figure 18.12 and Figure 18.10. In Figure 18.12 the evolution switches take place at the fold curve over the bifurcation set B, with a hysteresis in between them. After an evolution switch the new species is stable, and it remains stable even if the

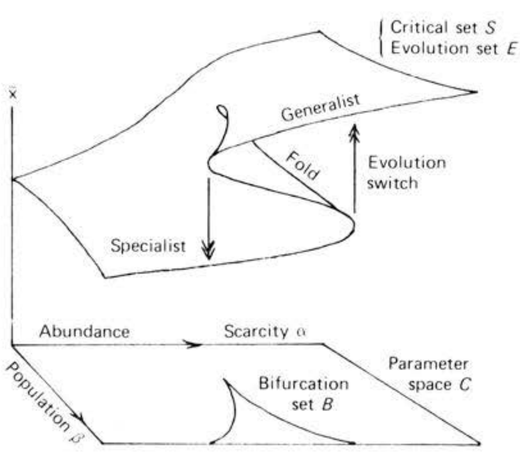

Figure 18.12. Specialists and generalists.

parameter is moved back again across the bifurcation set where the switch occurred. Therefore both types of beak can coexist stably together over the same parameter point in the interior of B. By contrast, in Figure 18.10 the decision switches take place over the Maxwell set M. There is no hysteresis, and therefore two decision makers cannot hold opposite decisions over the same parameter point. After a decision switch, the new decision is vulnerable to reversal if the parameter is moved back again across the Maxwell set, unless its stability is reinforced by the third ingredient of the global mechanism.

Example 4 : The Evolution of Roles in Society

The last example of biological evolution suggests a model for the social evolution of roles in society.

Let X be a space describing the roles. For example, if we were studying prehistoric hunter–gatherer roles (Reynolds and Zeigler 1979), we might take $X = \mathbb{R}^2$, with the first variable measuring the proportion of time spent in hunting and the second measuring that spent in gathering. If we were studying the present-day roles of men and women we might take $X = \mathbb{R}^2$ to measure the time spent in jobs and careers compared with that spent in domestic work. If we were interested in a more complicated combination of roles· we could take X multidimensional.

Let Y describe the possible future structures of society, including the size of population, the resources and technology, etc., and let P be the probability of such structures. Let $L(x, y)$ represent the unsuitability of role x for society y. For instance, we might invent some scale running from perfectly suited at $L = 0$ to totally unsuitable at $L = 1$. Alternatively, we might use L to measure the rewards offered to role x by society y. Then $R(x)$ will represent the unfitness of role x for the future society.

For example, in a hunter–gatherer study it could be that most of the individuals in the society did both jobs, and so the existing role would be presented by a single minimum of R near the point $(h, 1 - h)$, where h is the average proportion of time spent in hunting. However if the parameters of society changed, for instance if there were an increase in population or technology, then R might develop two minima, which would then drift toward the points $(1, 0)$ and $(0, 1)$. In other words a division of labor would occur. The individuals, particularly those growing up, would perceive that in the future it would probably be more rewarding to become either a specialist hunter or a specialist gatherer. In some societies the specialist roles became sex-linked.

An interesting point to notice is that although society may be changing gradually, or the population growing slowly, the resulting division of labor and separation of roles may occur suddenly. For instance, in Figure 18.12 it occurs at the threshold represented by the cusp point. Figure 18.12 also suggests that long periods of abundance, perhaps stimulated by new technology, may encourage the evolution of specialists to fill sociological niches and therefore favor the division of labor. Conversely, long periods of scarcity may have the opposite effect, encouraging the evolution of generalists and favoring self-sufficiency. Even today, during periods of prosperity specialists tend to thrive, but during recessions the adaptable survive better.

In a more general model the various local minima of R represent a variety of well-defined roles or sociological niches that a young person can aim for. Here the roles are in fact perceptions in the eyes of society, and the local mechanism that keeps a role at a minimum of R is natural selection, just as in biological evolution. If a gradual change in society causes a local minimum of R to move gradually, then society will perceive that a slight change in the role would make it fitter, and so the role will change accordingly. The role will respond continuously at the same time as keeping its name.

Meanwhile there is no global mechanism because different roles can coexist. The fact that the local minimum representing role B might be dropping until it is lower than that representing role A does not mean that role A will switch into role B. On the contrary, it is the *individual* who can decide to switch from A to B, while both *roles* continue to exist.

As well as the creation of roles we also observe their disappearance. The butcher and baker continue to thrive, but the candlestick-maker has been replaced by craftsmen, electricians, and manufacturers. This disappearance of roles gives a nice illustration of the difference between decision making and evolution. For individuals can decide to switch to a more rewarding role as soon as it becomes fitter, but the old role may survive in a anachronous form as long as the local minimum of R continues to exist. If the minimum ceases to exist (as in Figures 18.1 and 18.2), then the role will suddenly disappear. Thus society tends to be littered with dying anachronous roles existing alongside the living roles (where living means occupied by living individuals).

Of course different individuals will have different preferences. So far we have only considered a single loss function L representing the rewards offered by society as a whole, but each individual will have personal skills and preferences, and therefore will have a separate L for his or her own role. Consequently, a role that appears anachronous to society could appear fit to a particular individual and could even provide an unoccupied sociological niche. Thus society tends to be enriched by a few individuals keeping the anachronous roles alive. For instance, the hunter's role was probably jealously guarded by a few eccentrics long after their society had turned to agriculture, and no doubt the cherished weapons of those eccentrics were buried affectionately alongside them, providing the unsuspecting archaeologist with some anachronous data. Instances of anachronous roles can be observed amongst academics today, e.g., the old-fashioned archaeologist who refuses to believe in carbon dating.

In summary, the purpose of this example was to show how aspects of both decision making and biological evolution can appear in the same model of social evolution. Individuals can decide what role to adopt with their eyes open, but the roles themselves evolve blindly by natural selection, as in biological evolution.

Example 5 : The Evolution of Society

This example suggests the underlying data for the social evolution described in the Introduction.

Let X denote the possible social structures of a society. Let Y denote the possible future resources of the society, including not only the positive resources such as food, labor, energy, raw materials, technology, etc., but also the negative resources such as mouths to feed, bodies to house, territories to defend, and so forth. Given structure x, let $P(x, y)$ denote the probability of resources y. Here $P(x, y)$ may depend strongly on x, as illustrated by the different standards of living achieved by countries with comparable potential resources but different social structures in the world today.

Given resources y, let $L(x,y)$ measure the intolerance of society towards structure x. We invent some scale running from desirable at $L = 0$ to intolerable at $L = 1$. For example, if resources are scarce then society may not tolerate too great an imbalance in their distribution. If, however, there is a likely improvement of resources in the future, then society may be prepared to tighten its belt and accept the loss of a certain amount of individual freedom for a while in order to serve the needs of production. As resources improve, society may no longer tolerate such loss, and may demand less authoritarian a structure. And so on. Obviously L will depend on cultural parameters as well as the economic variables represented by y.

Then $R(x) = \int L(x,y)P(x,y)dy$ will measure the unfitness of structure x. As the resources P change, and as the cultural concept L of what is tolerable changes, so the unfitness R will change. The local mechanism keeping the structure at a local minimum of R is socioeconomic pressure, and the global mechanism for switching it to the fittest structure at the global minimum of R is government or revolution, as explained in the Introduction.

The main difference between this example and the last is the presence of the global mechanism, which eliminates the anachronous structures. In the last example the anachronous roles were harmless, because individuals could decide to switch out of them, and so society did not bother to invent a global mechanism to eliminate them. By contrast, individuals cannot escape an anachronous structure, and so they combine to devise a global mechanism to eliminate it.

Example 6: Political Parties

This example is a simpler version of the last, relating it to current politics. One of the problems facing a political party at the approach of an election is the decision as to what policy or what electoral image to present to the electorate. Suppose for the sake of simplicity that this decision can be represented by a point on the traditional political spectrum running from the Left to the Right. Take this spectrum to be the decision space X. The model also holds equally well for more complicated spaces of ideologies (Zeeman 1979).

Let Y represent the different types of voters in the electorate. Y may contain recognizably different components, and each component may contain a spectrum of types or be multidimensional. Let $P(y)$ measure the electoral strength of type y. Let $L(x, y)$ measure the lack of appeal of policy x to voter of type y. The $R(x)$ will measure the electoral weakness of policy x, and the local minima of R will represent electorally the strongest policies. The local mechanism is cooperation and realism, for politicians with similar ideologies will tend to cooperate together to present a common policy, and that policy will tend to be drawn to the nearest local minimum of R by the realism of having to appeal to the electorate. Meanwhile the global mechanism is the election and the constitution, for the election will select the lowest minimum, and the constitution will maintain that policy in power for a specified period.

The establishment of a particular social class, or the persistence of a particular economic problem, may cause the growth of an electorally strong component of Y, thus providing an electoral niche and stimulating the evolution of a political party and specialist policy to fill that niche. Conversely, the blurring of class differences may tend to skew P, and may cause the evolution of more generalist policies toward the center. The onset of short-term economic problems may have a similar effect and cause a party to be split between a specialist policy and a generalist policy. In other words if P_0 represents the party membership, then $R_0(x) = \int L(x,y)P_0(y)dy$ will be bimodal, with its two local minima representing the two policies. This can be represented by Figure 18.12, with the normal factor α representing the proportion of party membership outside the original traditional constituency, and the splitting factor β representing the onset of economic problems. Figure 18.13 illustrates various time paths in the parameter space.

Path 1 represents a gradual broadening of party membership followed by a gradual onset of economic problems, resulting in the gradual evolution from a specialist policy to a generalist policy. Path 2 represents a similar change, except that the events happen in the opposite order, so that the path crosses both sides of the cusp B as well as the Maxwell set M. For simplicity of exposition suppose we are discussing a party of the Left, so that initially R_0 has a unique minimum to the left of X. When path 2 crosses the left side of the cusp, another local minimum is created near the center of X, and a right wing of the party is formed in support of the corresponding alternative policy. When the Maxwell set M is crossed, the right wing achieves a majority in the party. There is then a conflict between local and global mechanisms, because the local mechanism tends to split the party, attracting each wing to its local minimum. Meanwhile the global mechanism is the party voting procedure and constitution, and if this

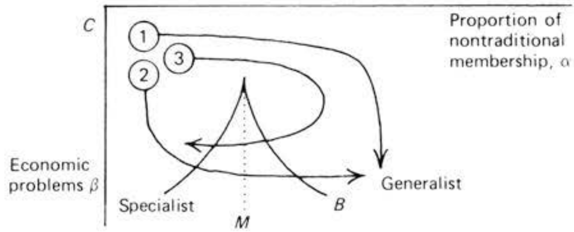

Figure 18.13. Time paths in the parameter space.

is strong enough it will unify the party at the lower minimum. Therefore it will cause a policy switch from left to center as the path crosses M. The old left policy will be retained as a possible alternative during the anachronous period, but will disappear when the path crosses the right side of B. Path 3 represents the opposite route: The onset of economic problems is followed by an increase of traditional membership, or equivalently a decrease in the nontraditional membership (attracted away, perhaps, to other parties). The result is the formation of a left wing as the path enters the cusp, a switch back to a traditional left-wing policy as the path crosses M, and the disappearance of the alternative center policy as it leaves the cusp.

Example 7: Doctors and Patients

In this last example we study the confusions that can arise when two decision makers use different loss functions and hence arrive at different risk functions. Call a decision maker *limited* if he or she only has local information and local utilities, so that a local minimum appears global. Call a decision maker *timid* if he only has a local mechanism, and no global mechanism. Both limited and timid decision makers will stay in a local minimum until it disappears, and so will appear to act blindly—like evolution—during an anachronous interval.

One can imagine many examples. For instance, without modern scientific tools some archaeological speculations were necessarily limited. To a government an electorate may appear limited in its lack of understanding of the need for unpopular measures. Conversely, to an electorate a government may appear timid in its reluctance to switch decisions. To illustrate the subtleties of the point we choose a domestic example of misunderstandings between a doctor and his patient.

There is often a delay, during which the symptoms of an illness are increasing before a person will acknowledge that he is ill, and likewise a delay before he acknowledges he is well again. The doctor has two problems: First, he may be curious to know what prompted the patient to decide that he was ill and what triggered the decision to come to the doctor for advice. Second, he may be frustrated when the recovering patient stubbornly persists in behaving as if he were still ill, when it would be to his advantage to behave as if he were well again. We apply the model to elucidate these problems.

Let X represent the possible treatments. For instance, if we measure the time spent on various strategies such as going to bed, visiting the doctor, going to the hospital, etc., then the treatment will be represented by a point x in $\mathbb{R}^n$. In particular, there is a point x_0 representing no treatment, i.e., normal life. Let y represent the possible symptoms of the illness. For instance, if we measure the intensity of each q different symptoms, each on some scale, then all the symptoms will be represented by a single point y in $\mathbb{R}^q$. Let $P(x,y)$ denote the probability of having symptoms y in the future as a result of treatment x today. Let

$$L_0(x) = \text{the cost to the patient of treatment } x;$$

$$L_1(y) = \text{the cost to the patient of symptoms } y.$$

We assume that the costs include not only financial costs, but also penalties for time lost, inconvenience, physical and psychological handicaps, etc., and that they can be measured compatibly so as to be added together to give a total cost $L(x,y) = L_0(x) + L_1(y)$. Consequently there are three risk functions:

$$\begin{aligned}
R_0(x) &= \textstyle\int L_0(x)P(x,y)dy = L_0(x) = \text{cost of treatment } x,\\
R_1(x) &= \textstyle\int L_1(y)P(x,y)dy = \text{medical risk of treatment } x,\\
R(x) &= \textstyle\int L(x,y)P(x,y)dy = R_0(x) + R_1(x) = L_0(x) + R_1(x)\\
&= \text{total risk of treatment } x.
\end{aligned}$$

We assume L_0 has a minimum at no treatment x_0, that is to say normal life. Suppose that x_1 represents the appropriate treatment for the illness, so that R_1 has a minimum at x_1. For instance in Figure 18.14 the point x_1 represents going to the hospital. Initially, before the patient becomes ill, the probability of no symptoms is 1, and the cost of no symptoms is 0, and so $R_1 = 0$. Therefore $R = R_0 = L_0$, with minimum at x_0, so the patient leads a normal life. Then as the symptoms appear, R_1 increases, causing R to follow an increasing sequence of graphs as in Figure 18.1 (remembering of course that X may be n-dimensional, whereas in Figures 18.1 and 18.14 it is drawn as only one-dimensional). The patient will decide to switch to treatment x_1 at the Maxwell point [graph number 4 in Figure 18.1(a)]. The timid patient will delay until the bifurcation point [graph number 6 in Figure 18.1(b)], although during the anachronous interval it would have been better for him to have sought treatment.

Let us look at the doctor's first problem—his curiosity as to what prompted the switch. From the medical point of view of treating this type of illness, as opposed to treating this particular patient, the doctor will be looking at the risk function R_1, with a unique minimum at x_1. Until he begins to treat this parti-

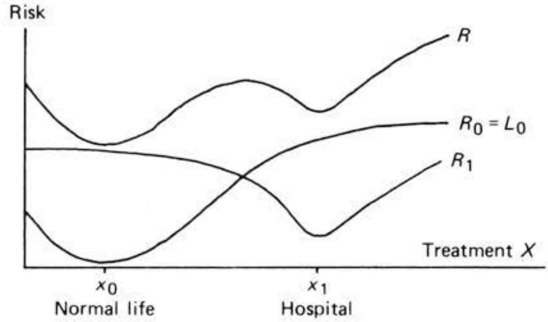

Figure 18.14. Three risk functions, $R = R_0 + R_1$.

cular patient he cannot know R_0, which is the cost to this patient of the treatment in terms of disruption to normal life, and so he must either ignore R_0, or include an estimated value for the hypothetical general patient. If he ignores R_0, then R_1 can only give him information on the best treatment x_1, and cannot give him any information about either of the switch points 4 or 6. Therefore, there can be no general answer as to what triggers the switch. Moreover from the point of view of R_1, even the intelligent decision switch 4 may appear as timid, while from the point of view of R prior to the decision switch the medical assessment R_1 must appear limited. Therefore, initially the patient may appear timed to the doctor, and the doctor may appear limited to the patient, because they are necessarily using different loss functions and therefore different risk functions.

We now turn to the doctor's second problem of frustration at the stubbornness of the recovering patient who persists in behaving as though he or she were still ill. The situation is represented by the graph of Figure 18.1 played in the reverse order 7, 6, . . . , 1. Since the doctor is by now treating the individual patient rather than thinking in terms of the general patient, he will have abandoned R_1 in favor of R, and be expecting the patient to make the decision switch back to normal life at the Maxwell point 4. However, by this time, the patient may have become an expert on his own illness. He may have become so accustomed to paying the fixed penalty $L_0(x_1)$, that he now discounts it by adopting the limited medical point of view of minimizing R_1 at x_1. Thus he is lead into persisting in the local minimum x_1 during the anachronous interval from 4 to 2 to the doctor's frustration. This time it is the patient who is limited, although to the doctor he again appears timid.

APPENDIX

Here is a definition of genericity sufficient for the classification theorems to hold. *Smooth* means that all partial derivatives exist to all orders. Let $X \subset \mathbb{R}^n$.

Let E be the ring of germs of smooth functions $\mathbb{R}^n, 0 \to \mathbb{R}$ stratified by the orbits of the group of germs of smooth diffeomorphisms $\mathbb{R}^n, 0 \to \mathbb{R}^n, 0$ acting on the right. Given $P{:}C \times X \times Y \to \mathbb{R}$ such that $\int P_c(x,y)dy = 1$, for all $c \in C$, $x \in X$, and given $L{:}C \times X \times Y \to \mathbb{R}$, define $R{:}C \times X \to \mathbb{R}$ by $R_c(x) = \int L_c(x,y)P_c(x,y)dy$. Call L *generic for* P if R is smooth and the map $R^+{:}C \times X \to E$ given by $R_c^+(x)(x') = R_c(x + x') - R_c(x)$ is transverse to the stratification. Suppose dimension $C \leq 5$.

THEOREM 1. If L *is generic for* P *then* L′ *is generic for* P′, *for sufficiently small perturbations* L′, P′ *of* L, P. *If* L *is not generic for* P *then there exist arbitrarily small perturbations* L′ *of* L *that are generic for* P.

In other words genericity is open dense. Therefore we are justified in using generic models.

THEOREM 2. If L *is generic for* P *then the resulting map* χ:S → C *is stable under sufficiently small perturbations of* L *and* P, *and all its singularities are elementary catastrophes.*

Therefore we are justified in using elementary catastrophes as models.

For proofs, see Zeeman (1977), together with the observation that any required perturbation R' of R can be obtained by choosing $L' = L - R + R'$.

ACKNOWLEDGMENTS

I am indebted for discussions to Maurice Dodson, Dermott Greene, Jeff Harrison, Colin Renfrew, Jim Smith, and Michael Thompson.

REFERENCES

Callahan, J., 1981, Bifurcation geometry of E_6, *Mathematical Modelling* 1, 283–309.

Callahan, J., 1982, A geometric model of anorexia and its treatment, *Behavioral Science.*

Dodson, M., 1975, Quantum evolution and the fold catastrophe, *Evolutionary Theory* 1, 107–118.

Harrison, P. J., and Smith, J. Q., 1979, Discontinuity, decision and conflict, in Bernado, J. M., DeGroot, M. H., Lindley, D. V., and Smith, A. F. M. (eds.), *Bayesian Statistics*, Proceedings First International Meeting, Valencia University Press, Spain.

Isnard, C. A., and Zeeman, E. C., 1976, Some models from catastrophe theory in the social sciences, in Collins, L. (ed.), *The Use of Models in the Social Sciences*, London, Tavistock Publications, 44–100.

Renfrew, C., 1978, The anatomy of innovation, *British Archaeological Reports International Series (Supplement)* 47, 89–117.

Renfrew, C., 1979, Systems collapse as social transformation: catastrophe and anastrophe in early state societies, in Renfrew, C., and Cooke, K. L. (eds.) *Transformations: Mathematical Approaches to Culture Change*, New York, Academic Press, 481–506.

Reynolds, R. G. D., and Zeigler, B. P., 1979, A formal mathematical model for the operation of consensus-based hunting–gathering bands, in Renfrew, C., and Cooke, K. L. (eds.), *Transformations: Mathematical Approaches to Culture Change*, New York, Academic Press, 405–418.

Smith, J. Q., 1977, *Problems in Bayesian Statistics Relating to Discontinuous Phenomena, Catastrophe Theory and Forecasting*, Doctoral dissertation, Warwick University.

Smith, J. Q., Harrison, P. J., and Zeeman, E. C., 1981, The analysis of some discontinuous decision processes, *European Journal of Operational Research*, 7, 30–43.

Thom, R., 1972, *Structural Stability and Morphogenesis*, English translation D. H. Fowler, 1975, New York, Benjamin.

Thompson, M., 1981, A three-dimensional model, in Douglas, M., and Ostrander, D. (eds.), *Essays in the Sociology of Perception*, London, Routledge and Kegan Paul.

Zeeman, E. C., 1977, *Catastrophe Theory: Selected Papers 1972–1977*, Reading, Mass., Addison–Wesley.

Zeeman, E. C., 1979, A geometrical model of ideologies, in Renfrew, C., and Cooke, K. L. (eds.), *Transformations: Mathematical Approaches to Culture Change*. New York, Academic Press, 463–479.

Zeeman, E. C., 1980, Catastrophe models in administration, *Association for Institutional Research, Annual Forum Proceedings* 3, 9–24.