European Journal of Operational Research 36 (1988) 379–392
North-Holland

Theory and Methodology

Divergence of choices despite similarity of characteristics: An application of catastrophe theory

Giacomo BONANNO
Department of Economics, University of California, Davis, CA 95616, USA

E. Christopher ZEEMAN
Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK

Abstract: In this paper we use catastrophe theory to analyse situations in which agents with similar characteristics and objectives and facing identical or similar environments make choices which are considerably different. We first provide two simple analytical examples of this phenomenon and then set up a general framework to which we apply the classification theorem of catastrophe theory.

Keywords: Catastrophe theory, decision theory, optimization, probability, expected utility

1. Introduction

Consider a set of agents, all of whom face the same environment (e.g. the same data or information) and the same set of choices. If the agents differ among themselves, we would expect them to make different choices. Intuition, however, suggests that similar agents with similar objectives should make similar choices. The purpose of this paper is to show that it is possible to have divergence of choices despite similarity of characteristics and objectives, and, furthermore, that this is a stable phenomenon, in the sense that it cannot be eliminated by means of small changes in the specification of the model. A consequence of this is that such phenomena cannot be considered 'unlikely' or 'pathological'.

The mathematical tool which we use is catastrophe theory [1]. In the following two sections we illustrate the phenomenon with the help of two simple analytical examples, while in Section 4 we set up a general framework and apply the classification theorem of catastrophe theory to it. Section 5 provides a summary.

2. Example 1: Forecasting the value of a random variable [2]

Consider a situation in which a number of agents are asked to make a forecast x about the

We are grateful to J.T. Dockey and two anonymous referees for helpful comments and suggestions. A first version of this paper was presented at the European Meeting of the Econometric Society (Budapest, September 1986).

Received March 1987; revised August 1987

[1] For an elementary exposition of catastrophe theory see Chillingworth (1976) and Fararo (1978).

[2] This example is based on Smith, Harrison and Zeeman (1981) and Zeeman (1982). It will be pointed out later that the work of Smith (1978) is also directly relevant to this example.

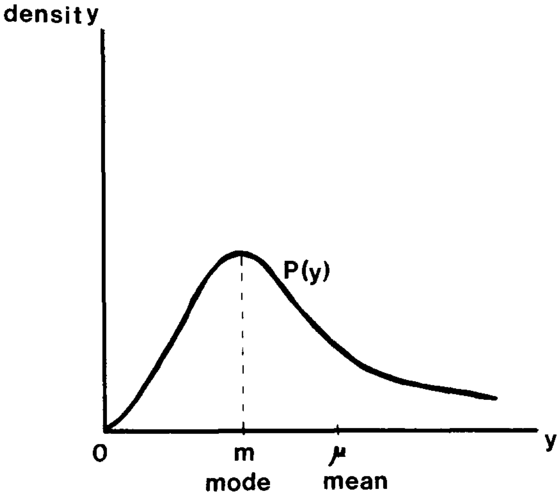


Figure 1. Skew probability density function

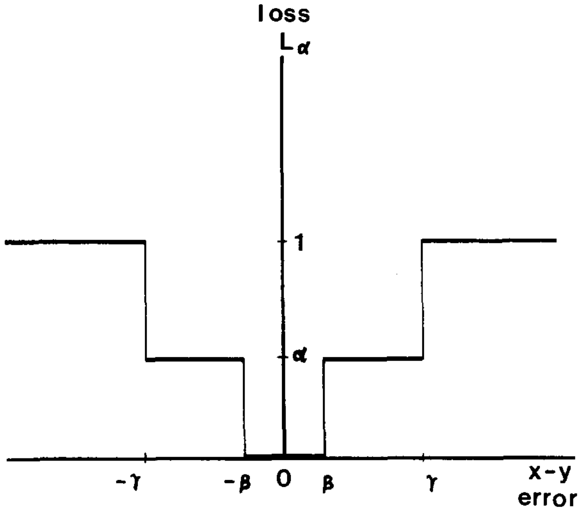


Figure 2. Two-step loss function

value of a random variable Y (for example, the price of a particular stock). The probability distribution of Y is common knowledge and given by the density function $P(y)$. Let x be the forecast value and y the observed value. Then the payoff is as follows: if the forecast is approximately correct, i.e. if $0 \leqslant |x-y| \beta$ (where $\beta > 0$ is small), the agent receives a prize $p > 0$; if the forecast is considerably wrong, i.e. if $|x-y| > \gamma$ (where $\gamma > 0$ is large) the agent has to pay a fine $s > 0$; in every other case the agent receives and pays nothing.

Suppose the density function $P(y)$ is as shown in Figure 1.

We have deliberately chosen a skew distribution so that the mode m is different from the mean μ. The mode represents the most likely outcome, while the mean represents the expected outcome. Will an agent's forecast be based on the mean or on the mode?

We consider a large number (in fact a continuum) of agents. Each agent is identified by a value, between 0 and 1, of the parameter w, which gives the weight the agent attaches to the prize p (and $(1-w)$ is the weight he attaches to the fine s). Intuitively we would expect the agent for whom the prize is all that matters ($w=1$) to follow the mode m (the most likely outcome), and the agent who is entirely worried about the fine ($w=0$) to follow the mean μ. We would also expect that as w increases from 0 to 1 the forecast varies continuously from the mean to the mode. We now show that this need not be the case.

A agent $w \in [0, 1]$ will view the game as having the following loss function as payoff

$$L_w(x, y) = \begin{cases} -wp & \text{if } 0 \leqslant |x-y| \leqslant \beta, \\ 0 & \text{if } \beta < |x-y| \leqslant \gamma, \\ (1-w)s & \text{if } |x-y| > \gamma. \end{cases} \qquad (1)$$

For each agent w, define the *risk* of making forecast x as follows:

$$R_w(x) = \int L_w(x, y) P(y) \, \mathrm{d}y. \qquad (2)$$

In other words, the risk function is the expected loss. We assume that each agent makes the forecast x^* that minimizes $R_w(x)$.

Let

$$\alpha = wp/[wp + (1-w)s]. \qquad (3)$$

Then we can normalize L_w by defining

$$L_\alpha(x, y) = \frac{wp + L_w(x, y)}{wp + (1-w)s}$$
$$= \begin{cases} 0 & \text{if } 0 \leqslant |x-y| \leqslant \beta, \\ \alpha & \text{if } \beta < |x-y| \leqslant \gamma, \\ 1 & \text{if } |x-y| > \gamma. \end{cases} \qquad (4)$$

Such a normalization is admissible because affine [3] changes of L induce the same affine changes in R and so do not alter the critical points of R. The

[3] A transformation T: $\mathbb{R} \to \mathbb{R}$ is called *affine* if it takes the form $T(x) = ax + b$, where a and b are constants.

two-step loss function L_α is shown in Figure 2.

As w varies between 0 and 1, so does α and we can identify an agent with a value $\alpha \in [0, 1]$. The parameter α can be interpreted as a measure of *confidence*: the higher α, the more confident the agent.

The risk function

$$R_\alpha(x) = \int L_\alpha(x, y) P(y)\, \mathrm{d}y \tag{5}$$

can now be computed as follows. Regard L_α as the constant function, with value 1, minus two rectangles, the first of width 2β and height α, and the second of width 2γ and height $(1-\alpha)$. Therefore

$$\begin{aligned} R_\alpha(x) &= \int L_\alpha(x, y) P(y)\, \mathrm{d}y \\ &= 1 - \alpha B(x) - (1-\alpha)\Gamma(x) \end{aligned} \tag{6}$$

where

$$B(x) = \int_{|x-y| \leqslant \beta} P(y)\, \mathrm{d}y = \int_{x-\beta}^{x+\beta} P(y)\, \mathrm{d}y, \tag{7}$$

$$\Gamma(x) = \int_{|x-y| \leqslant \gamma} P(y)\, \mathrm{d}y = \int_{x-\gamma}^{x+\gamma} P(y)\, \mathrm{d}y. \tag{8}$$

It is interesting to note that the risk function is smooth, despite the fact that the loss function is not continuous (it is a step function). The critical points of B are given by

$$\mathrm{d}B/\mathrm{d}x = P(x+\beta) - P(x-\beta) = 0. \tag{9}$$

Therefore B has a unique maximum at m', say, where m' is the mid-point of the unique horizontal chord of P of length 2β. Since β is small, m' is near the mode of P. Similarly, Γ has a unique maximum at μ', say, where μ' is the mid-point of the chord of length 2γ. If γ is of the order of about twice the standard deviation of P then it can be seen from Figure 1 that μ' is near the mean μ of P, as shown in Figure 3.

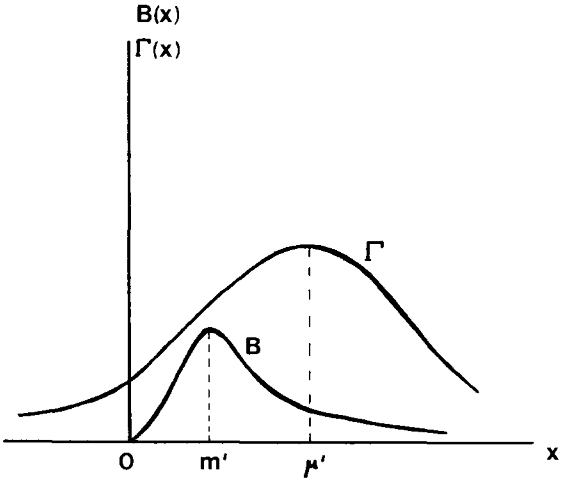


Figure 3. The functions $B(x)$ and $\Gamma(x)$

Therefore if $\alpha = 0$ then R_α has a unique minimum at μ' and if $\alpha = 1$ then R_α has a unique minimum at m'. If $0 < \alpha < 1$ then R_α is a linear combination of B and Γ, and so it is either unimodal, with a unique minimum between m' and μ', or else bimodal, with two minima, one near the mode and the other near the mean, as shown in Figure 4. As the parameter α varies from 0 to 1, the decreasing family of loss functions give rise to a smooth decreasing family of smooth risk functions, one possibility being the one shown in Figure 4 (in the Appendix we give a very simple example of a bimodal risk function arising from a unimodal probability density function and the two-step loss function (4) illustrating this possibility).

Let $x^* = g(\alpha)$ be the absolute minimum of $R_\alpha(x)$. For the family shown in Figure 4 there is a critical value α^* of the parameter for which the risk function $R_\alpha(x)$ has both minima at the same level. If $\alpha < \alpha^*$ then the lower minimum $g(\alpha)$ will be a point μ' near the mean, and if $\alpha > \alpha^*$ then

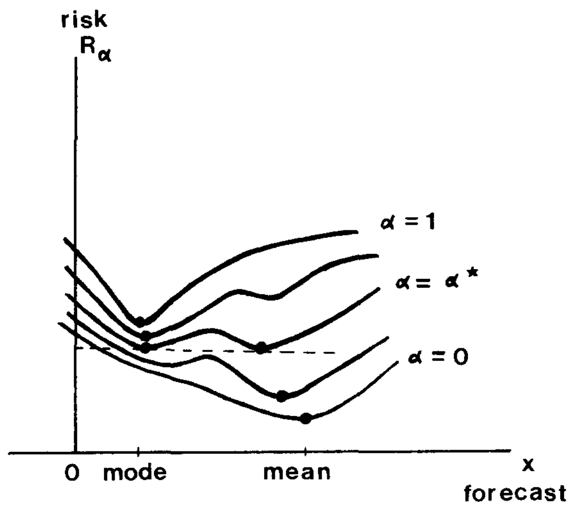


Figure 4. The family of risk functions. The absolute minimum of each risk function is indicated by a solid dot. At the critical value α^* the risk function has both minima at the same level, indicated by the horizontal dotted line

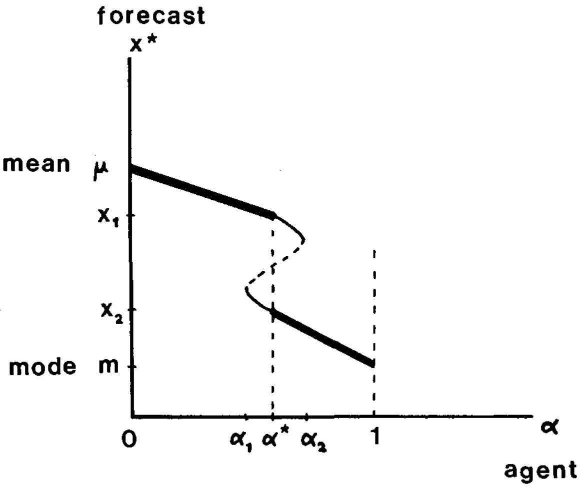


Figure 5. Discontinuous choice function

g(α) will be a point m' near the mode. Therefore if α increases past α^* then $g(\alpha)$ will switch discontinuously from μ' to m'.

This is shown more clearly in Figure 5. The smooth S-shaped curve consists of all points (α, x) for which x is a stationary value (minimum or maximum) of R_α(x). The thick part of the curve represents the absolute minima, the thin part represents the local (but not absolute) minima, while the dotted part represents maxima. Therefore the graph of $x^* = g(\alpha)$ (absolute minima) is the thick curve, with the discontinuity at α^*. We call this function $x^* = g(\alpha)$ the *choice function*, since it gives the forecast chosen by agent α (choice functions will be discussed in more general terms in Section 4).

We therefore observe a phenomenon of *polarization*, with some agents making forecasts near the mean and the remaining agents making forecasts near the mode, while no agent makes a forecast in between. Figure 5 also illustrates two phenomena —divergence and inaccessibility—which will be discussed at greater length in Section 4.

Values of x between x_1 and x_2 are not forecast by any agents (*inaccessibility*; these are values which correspond to local maxima or local—but not global—minima of R_α) and agents with similar characteristics and objectives (e.g. agents α_1 and α_2) end up making forecasts which are considerably different (*divergence*).

In the preceding discussion the density function P was fixed, while the loss function varied within a one-parameter family, L_α. We can now extend the analysis by considering a one-parameter family of density functions P_η. We define the parameter η as follows

$$\eta = \mu - m = \text{mean minus mode}. \tag{10}$$

We can interpret η as a measure of the *ambiguity* of the information about the random variable Y. A symmetrical distribution like the normal would have $\eta = 0$. If $\eta > 0$ then P is skewed to the right as in Figure 1, and if $\eta < 0$ then P is skewed to the left. First consider only non-negative values of η. Let $x^* = g(\alpha, \eta)$ be the forecast chosen by agent α when he faces the density function P_η. If $\eta = 0$ the mode and the mean coincide and therefore all agents will make the same forecast $x = \mu = m$. Therefore $g(\alpha, 0)$ is a horizontal straight line, hence continuous. If η is positive we are in the situation analyzed previously where the function $g(\alpha, \eta)$ has a discontinuity as shown in Figure 5. The classification theorem of catastrophe theory, which will be stated in Section 4, enables us to conclude that the graph G of the choice function $g(\alpha, \eta)$ is a surface which is equivalent to that shown in Figure 13(c) (cusp). If the information is skewed the other way ($\eta < 0$), implying $\mu < m$, then another symmetrically placed cusp appears, with the orientation reversed. The graph G of $g(\alpha, \eta)$ would therefore look like Figure 6. If $\eta = 0$ our two agents α_1 and α_2 make the same forecast, but if P becomes skewed either way then they will find their forecasts diverging in opposite directions, with α_1 always following the mean and α_2 following the mode.

We conclude this section with a remark. We have considered the case where a two-step loss

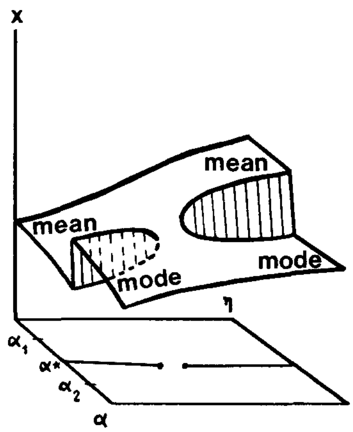


Figure 6. Skewing either way gives two cusps

function coupled with a unimodal density function gives rise to a bimodal risk function (an example is given in the Appendix) and therefore to the possibility of a discontinuous choice function. It is worth stressing that in general the risk function need not be bimodal. Smith (1978) gives sufficient conditions on the skew density which, together with a double-step loss function, ensure bimodality of the resultant risk function (see also Smith, Harrison and Zeeman, 1981) [4].

3. Example 2: Asking for the 'right' salary [5]

Consider the case of a firm and a large number of workers. The firm believes that each worker's productivity is the same and equal to S, which, without loss of generality, we can take to be a positive number less than or equal to 1. The firm, however, does not want to disclose its beliefs; instead, it asks each worker to state the salary $x \in [0, 1]$ at which he is willing to work and it will employ the worker if and only if $x < S$. All workers have the same utility function [6]

$$U(x) = x. \tag{11}$$

They differ, however, in their beliefs concerning the value of S. A worker's beliefs are expressed by a density function h: $[0, 1] \to \mathbb{R}$ whose cumulative distribution function we denote by H. Thus $H(x)$ is the probability, according to the worker's beliefs, that $S \leqslant x$, that is, that if he requests salary x, the firm will not employ him. In other words, the worker attaches probability $1 - H(x)$ to the event that the firm will employ him if he requests salary x.

We want to consider a one-parameter family of beliefs, H_α, where α is a one-dimensional parameter and each value of the parameter identifies a worker (or agent); furthermore, we want increasing α to mean increasing pessimism. We say that worker α is *more pessimistic* than worker α' if his beliefs, H_α, dominate those of worker α', $H_{\alpha'}$, in the sense of *strict first-order stochastic dominance*, that is, if

$$H_\alpha(x) \geqslant H_{\alpha'}(x) \quad \text{for all} \quad x \in [0, 1] \text{ and } H_\alpha \neq H_{\alpha'} \tag{12}$$

(recall that $H(x)$ is the probability, according to the worker's beliefs, that if he requests salary x the firm will not employ him) [7].

For the time being we shall leave aside the parameter α (worker) and consider the following two-parameter family of cumulative distribution functions (beliefs):

$$H_{b,c}(x) = \begin{cases} bx & \text{if } 0 \leqslant x \leqslant (1-c)/(b-c), \\ 1 - c + cx & \text{if } (1-c)/(b-c) \leqslant x \leqslant 1. \end{cases} \tag{13}$$

Define

$$P_{b,c}(x) = 1 - H_{b,c}(x). \tag{14}$$

For a worker whose beliefs are represented by the point (b, c), $P_{b,c}(x)$ gives the probability that $x < S$, that is, the probability that if he asks for salary x he will be employed at that salary. From (13) we obtain

$$P_{b,c}(x) = \begin{cases} 1 - bx & \text{if } 0 \leqslant x \leqslant (1-c)/(b-c), \\ c - cx & \text{if } (1-c)/(b-c) \leqslant x \leqslant 1. \end{cases} \tag{15}$$

Figure 7 illustrates the function $P_{b,c}(x)$ for the cases where $b = c = 1$ and $0 < c < 1 < b$. According to our definition (b, c) represents more pessimistic beliefs than (b', c') if and only if $P_{b,c}(x) \leqslant$

[4] The class of skew densities which—together with the two-step loss function (4)—give rise to a bimodal risk function includes the lognormal, inverse gamma and the Pareto distributions (cf. Smith, 1978), as well as the one constructed in the Appendix. It does not, however, include commonly used densities, such as the gamma.

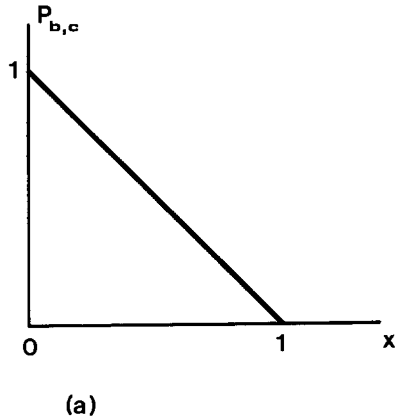
[5] The example of this section is formally similar to Bonanno (1987). In fact, the function $f_{b,c}(x)$ given by (19) is formally identical to a revenue function.

[6] The utility function was chosen to be linear (risk-neutrality) only in order to simplify the analysis, but similar results can be proved with concave utility functions (risk-aversion) (cf. Bonanno, 1986).

[7] The intuition behind our definition of greater pessimism is reinforced by the following well-known property (cf. Lippman and McCall, 1982, pp. 215–216): H_α dominates $H_{\alpha'}$ in the sense of strict first-order stochastic dominance if and only if for every increasing function U,

$$\int U(x) h_\alpha(x)\, dx < \int U(x) h_{\alpha'}(x)\, dx.$$

Therefore, worker α is more pessimistic than worker α' if and only if his expected utility is less than that of worker α' (assuming that they have the same utility function).



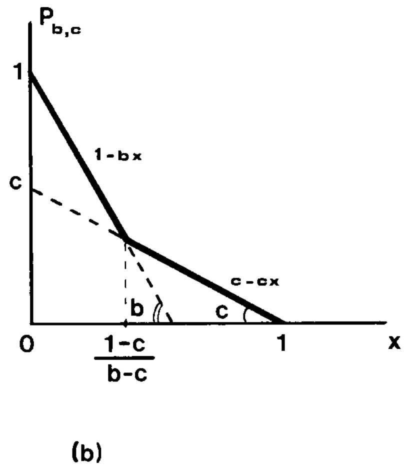


Figure 7. The function $P_{b,c}(x)$. We have labelled the angles by the value of their tangents. (a) Case $b=c=1$; (b) Case $0<c<1<b$

$P_{b',c'}(x)$ (for all x, with strict inequality for some x). It is easy to check that a sufficient condition for this is

$$b \geqslant b' \text{ and } c \leqslant c' \text{ and not both equal.} \tag{16}$$

The two-parameter family of c.d.f.'s given by (13) arises from the following two-parameter family of two-step density functions, illustrated in Figure 8:

$$h_{b,c}(S)=\begin{cases} b & \text{if } 0 \leqslant S \leqslant (1-c)/(b-c), \\ c & \text{if } (1-c)/(b-c) \leqslant S \leqslant 1. \end{cases} \tag{17}$$

By (16), given any point in the parameter space, a movement in the South, East or South-East direction is associated with increasing pessimism.

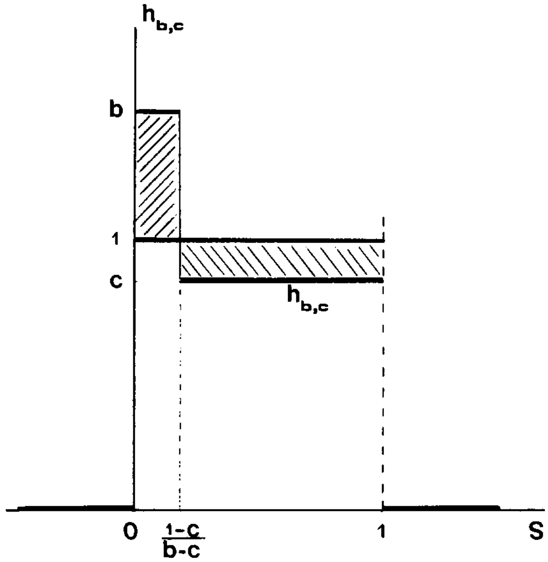


Figure 8. The density function $h_{b,c}(S)$. The two shaded areas are equal

We shall restrict our attention to values of the parameters satisfying the condition

$$0 \leqslant c < 1 < b. \tag{18}$$

The shaded area in Figure 9 illustrates the region defined by (18) and the arrows at point V indicate the directions associated with increasing pessimism (East, South or South-East).

Let

$$f_{b,c}(x)=U(x)P_{b,c}(x)=xP_{b,c}(x). \tag{19}$$

Thus $f_{b,c}(x)$ is the expected utility of asking for salary x. It follows from (15) that the function

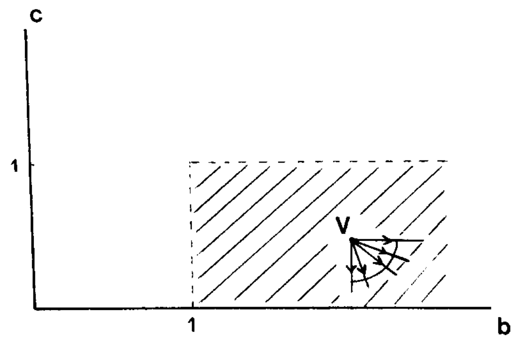


Figure 9. The arrows denote the directions which are associated with increasing pessimism (from any given starting point V)

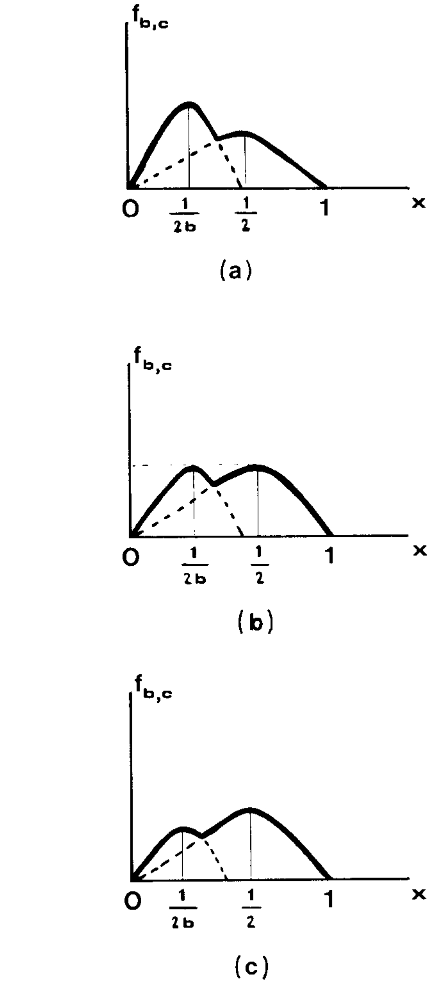


Figure 10. The function $f_{b,c}(x)$ (a) Case $1/4b > c/4$; (b) Case $1/4b = c/4$; (c) Case $1/4b < c/4$

$f_{b,c}(x)$ is the union of two parabolas, as illustrated in Figure 10 [8].

We shall assume that each worker asks for that salary x^* which maximizes $f_{b,c}(x)$. Let $x^* = g(b, c)$ be the point at which $f_{b,c}(x)$ reaches its maximum; then it is easy to check that

$$x^* = \begin{cases} 1/2 & \text{if } c > 1/b, \\ 1/(2b) & \text{if } c < 1/b, \end{cases} \tag{20}$$

while for $c = 1/b$ the function (19) has two global

[8] The case $1/(4b) > c/4$ (Figure 10(a)) includes two more cases, where the maximum of the parabola on the right goes inside the other parabola and therefore the function f becomes unimodal. Similarly for the case $1/(4b) < c/4$ (Figure 10(c)).

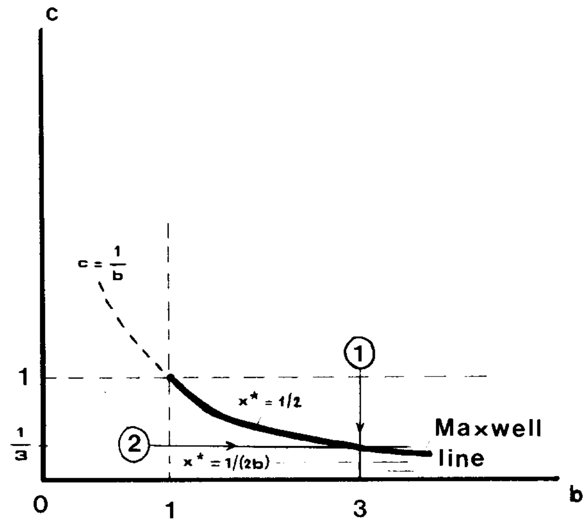


Figure 11. The line of equation $c = 1/b$ is the Maxwell line

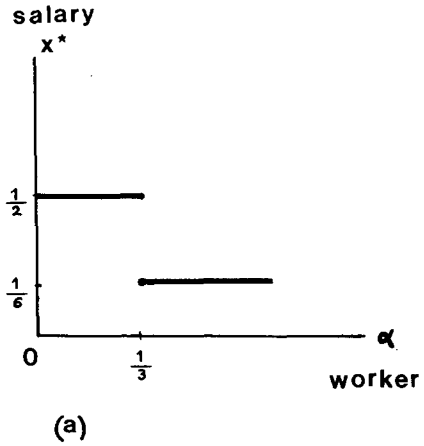
maxima at $x^* = 1/2$ and $x^* = 1/(2b)$ (cf. Figure 10(b)).

The above results are visualized in Figure 11. Any path in the (b, c)-space ϕ: $A \subset \mathbb{R} \to \mathbb{R}^2$ such that as α increases the corresponding point $\phi(\alpha)$ moves in the South, East or South-East direction (like paths (1) and (2) in Figure 11), can now be taken as the one-parameter family of beliefs referred to above: each value of α identifies a worker and increasing α means increasing pessimism. Whenever such path crosses the line (which, as we shall explain in Section 4, is called the *Maxwell line*) defined by the equation

$$c = 1/b \tag{21}$$

there will be a discontinuous jump in the salary $x^* = g(\alpha)$ requested by worker α (as in the previous section, we shall call the function $x^* = g(\alpha)$ the choice function). Path (1) in Figure 11 (given by $b = 3$ and $\alpha = 1 - c$) gives rise to a counterintuitive situation, as small differences in beliefs give rise to large differences in choices; however, the jump occurs in the 'right' direction: a more pessimistic worker asks for a lower salary (see Figure 12a). More surprising is the situation illustrated in Figure 12b, corresponding to path (2) in Figure 11 (given by $c = 1/3$ and $\alpha = b$). Here not only do we have a discontinuity, but the jump occurs in the 'wrong' direction: a more pessimistic worker asks for a much higher salary than his less pessimistic colleague!

The intuition behind these results is as follows.



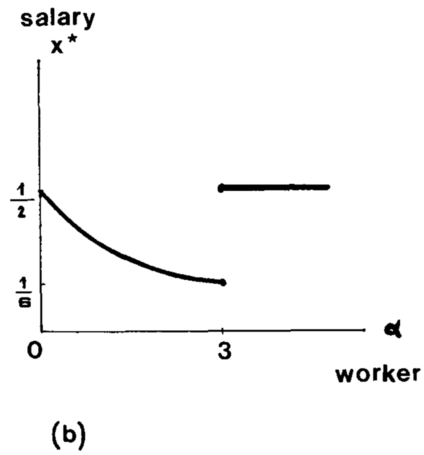


Figure 12. Evolution of requested salary along paths 1 and 2 (Figure 11) (a) Path (1) of Figure 11; (b) Path (2) of Figure 11

When b is close to 1, the probability function (15) is almost linear (cf. Figure 7(a)) and as a consequence the function f (given by (19)) has a unique maximum. With c fixed, an increase in b means that the probability of S being close to zero increases and the probability of intermediate values of S decreases, while the probability of S being close to 1 remains unchanged. The maximum of the function f, therefore, will move to the left towards 0, but at the same time a new local maximum will appear near 1. When the probability of S being close to zero becomes very high (that is, when beliefs become very pessimistic) then the worker knows that if he asks for a very low salary he is very likely to get the job, but his utility will also be very low. Therefore he may as well take a chance and ask for a high salary, since he believes there is a positive, although very small, probability that S is in fact large. Between a very likely life of misery and a not impossible life of luxury some workers will opt for the former and some—the slightly more pessimistic ones—for the latter. Here we can notice again the two phenomena of divergence and inaccessibility, which were pointed out in the previous section and will be discussed at length in Section 4.

We shall conclude this section with a remark. The functions which we considered in this example are continuous but not smooth: the functions (15) have a kink at the point $(1-c)/(b-c)$ and therefore the density functions (17) have a discontinuity at that point. It is clear from Figure 10, however, that the kink (and corresponding discontinuity) are irrelevant from our point of view, since the point $(1-c)/(b-c)$ is a (kinked) minimum of the functions $f_{b,c}(x)$ while we are interested in the maxima of those functions, which are smooth. We chose the family of functions (13) because of its simplicity. However, since smooth functions are dense in the space of continuous functions [9], we can choose a smooth approximation of the family (13), (thereby eliminating the discontinuity in the density functions $h_{b,c}$), and invoke the classification theorem of catastrophe theory (cf. Section 4) to conclude that any sufficiently close smooth approximation of the functions considered would exhibit the qualitative properties illustrated above.

4. A general framework

The purpose of this section is to set up a general framework which can accomodate the examples given above and enables us to state some general results.

Let A be a set of agents, with a topology on it which enables us to say whether two agents are similar (close) or different. In general each agent can be identified with a vector of k characteristics and therefore we can think of A as a subset of $\mathbb{R}^k$. We shall denote an element of A by α.

The state of the *environment* may affect the action chosen by each agent and we assume that we can measure the data or information about the

[9] Cf. Hirsh (1976, p. 47, Theorem 2.4).

environment by q parameters. Let E denote the space of possible parameter values, which will therefore be a subset of $\mathbb{R}^q$. Let an element of E be denoted by η.

Let X be the set of choices facing each agent. In general X will be a subset of $\mathbb{R}^n$.

We shall make two hypotheses.

Hypothesis 1. Each agent $\alpha \in A$ in each environment $\eta \in E$ has an *objective function*

$$f_{\alpha,\eta}\colon X \to \mathbb{R} \tag{22}$$

and the agent chooses x so as to maximize $f_{\alpha,\eta}$.

Hypothesis 2. Similar agents in similar environments have similar objective functions. This can be formalized by requiring

$$F\colon A \times E \times X \to \mathbb{R} \tag{23}$$

given by $F(\alpha, \eta, x) = f_{\alpha,\eta}(x)$ to be smooth [10].

Now it is a trivial but important consequence of these two hypotheses that although F is smooth, the resulting choice may *not* necessarily depend smoothly on the agent and environment, and can exhibit standard types of discontinuity. To focus attention upon this crucial fact we introduce the notion of the *choice function*

$$g\colon A \times E \to X \tag{24}$$

where $g(\alpha, \eta)$ is defined to be the choice made by agent α in environment η. In general $f_{\alpha,\eta}$ will have a unique global maximum at a unique point $x \in X$, and so $g(\alpha, \eta) = x$. However, in special cases $f_{\alpha,\eta}$ may have two global maxima at the same level at two different points x_1, x_2 and in this case $g(\alpha, \eta)$ will be double-valued and equal to the point pair $\{x_1, x_2\}$. Moreover, such special cases may be unavoidable if $A \times E$ is at least one-dimensional, because perturbations one way may raise one of the two maxima to be the unique global maximum, while perturbations the other way may raise the other maximum, causing a discontinuity in the choice function, as shown in Figure 13(b).

If $A \times E$ is two-dimensional, a further complexity can arise with three global maxima at the same level, as shown in Figure 13(d), but this is the worst possible case, as indicated by the theorem below.

Let G denote the graph of g, that is,

$$G = \{(\alpha, \eta, x) \in A \times E \times X \mid x = g(\alpha, \eta)\}. \tag{25}$$

We call G the *choice graph*. Over most points of $A \times E$ the graph G will be single-valued and continuous, but over certain points G can be multi-valued and discontinuous. The theorem below classifies the types of discontinuity that can arise.

We specialize to the case $k = q = 1$, that is, where agents are distinguished by a one-dimensional characteristic and face an environment that can be represented by a one-dimensional parameter. Note, however, that we impose no restrictions on n, the dimension of the set of choices X.

Let Ω be the space [11] of smooth functions F: $A \times E \times X \to \mathbb{R}$. We can now state the theorem, which is an immediate deduction from the deep classification theorem of elementary catastrophe theory due to René Thom and the trivial Gibbs phase rule (see Thom, 1972; Zeeman, 1977).

Theorem. *There exists an open dense subset Z of Ω, such that if $F \in Z$, then the resulting choice-graph G is a two-dimensional surface, which is locally equivalent at each point to one of the graphs shown in Figure 13. Furthermore, each graph in Figure 13 is stable in the sense that it cannot be eliminated by small perturbations of F.*

We remark that Thom's theorem classifies the types of smooth surfaces M of *stationary values* that can occur, whereas we are only interested in those stationary values that happen to be absolute minima. Therefore our choice graph G is a subset of M and in fact is a surface-with-boundary, with the boundaries occurring wherever the choice function is discontinuous. In Figure 13 the discontinuities are indicated by vertical lines (which are not actually part of G). The situation is the two-dimensional analogue of the one-dimensional graph shown in Figure 5: there the S-shaped curve is the smooth curve M of stationary values, and

[10] The smoothness assumption is not a strong one, since every continuous function can be approximated by a smooth function (cf. Hirsh, 1976, p. 47, Theorem 2.4).

[11] The topology of Ω is an obvious one: two functions are close if their values are close and their partial derivatives up to some order are close, and, to avoid problems at infinity, the closeness may tail off towards infinity. This is called the Whitney topology (see Zeeman, 1977).

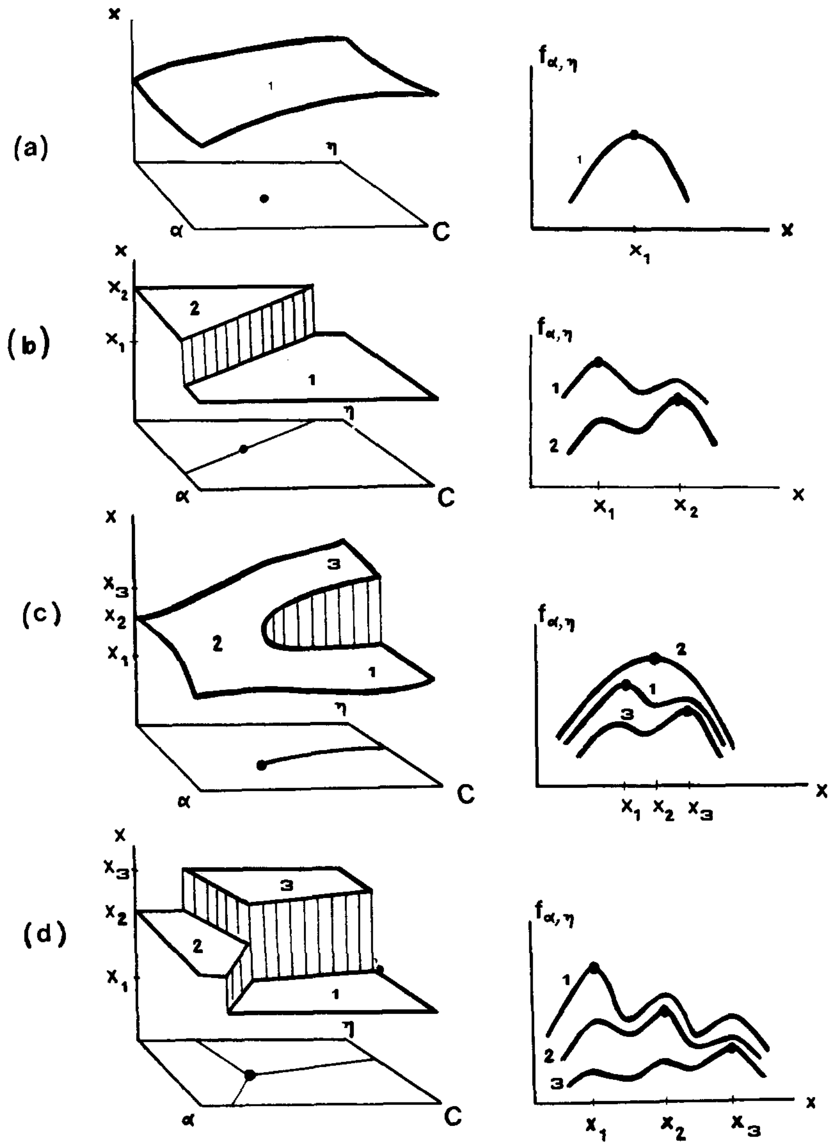


Figure 13. (a) Continuous choice; (b) Maxwell line; (c) Cusp point; (d) Maxwell point

the graph G is the subset given by the thick curve, which is in fact a curve-with-boundary, the boundary points occurring where the choice function is discontinuous.

Notice also that G lies in $A \times E \times X$ which is $(2+n)$-dimensional, while the graphs illustrated in Figure 13 lie in three dimensions. The difference is allowed for by the definition of 'local equivalence', which means that for each point $p \in A \times E$ there is a neighbourhood N of p in $A \times E$, a picture Q in Figure 13, and a diffeomorphism of N onto the horizontal square C of Q, throwing p onto the dot, and underlying a projection of $N \times X$ into $C \times (x\text{-axis})$ that throws the subset of G above N onto the graph in Q.

Case (a) of Figure 13 is the intuitive situation that one would expect to observe: choices vary continuously with characteristics and environment, and therefore similar agents in similar environments make similar choices.

Case (b) of Figure 13 is the counterintuitive situation of 'unavoidable' polarization: despite the fact that agents' characteristics are spread over a continuous range, we observe, essentially, only two types of choice rather than a continuum of choices. These two types of choice are separated by a line (a curve) in the (α, η)-plane which is called the *Maxwell line*. The Maxwell line is the set of points (α, η) at which the corresponding objective function $f_{\alpha,\eta}(x)$ has two global maxima. Two agents lying on either side of the Maxwell line may have characteristics so close as to be almost indistinguishable, yet will make very different choices, x_1 and x_2. In a given environment the set of agents is split into two by the Maxwell line. Similarly, for a given agent the set of environments is split into two by the Maxwell line. If the environment is gradually changing then the agent will suddenly switch decision as she crosses the Maxwell line. Different agents will switch at different times, so that the switch of decision will proceed like a wave along the spectrum of agents.

Case (c) of Figure 13 represents the 'threshold of polarization'. Here the Maxwell line starts at a point, which marks the onset of polarization in a gradually changing environment. The graph arises from the cusp catastrophe (see Thom, 1972; Zeeman, 1977). Before the threshold the agents face a continuous spectrum of choice, but after the threshold they are split into two classes facing essentially only two types of choice, x_1 or x_3. The middle choice x_2 is no longer accessible to them: we call this phenomenon *inaccessibility*. Again, agents who are close but pass the threshold on either side of the Maxwell line find themselves gradually diverging in their choice, although previously their choices had been relatively close: we call this phenomenon *divergence*. If the Maxwell line is at an angle to the environment axis, then there will be some agents who begin to diverge one way and then suddenly switch the other way (a very human trait!).

Case (d) of Figure 13 respresents a 'compromise' situation. The Maxwell line is Y-shaped, and at the vertex of the Y the three regions representing essentially three different choices meet.

We remark that this finite classification of types of discontinuity can be extended to higher dimensions, more precisely for parameter spaces $A \times E$ of up to five dimensions. The reason for this is that Thom's theorem classifying elementary catastrophes extends up to this dimension (see Zeeman, 1977).

We can conclude this section by showing how the examples of Sections 2 and 3 fit this general framework.

In the example of Section 2 the set of agents A is represented by the interval [0, 1] (cf. (3)) and the set of environments by the real line (the parameter η defined in (10) can take any value). The set of choices X is also the real line (each agent makes a one-dimensional choice consisting in forecasting the value of the random variable Y). Finally the objective function (22) is given by

$$f_{\alpha,\eta}(x) = -R_{\alpha,\eta}(x) = -\int L_\alpha(x, y) P_\eta(y)\, \mathrm{d}y. \tag{26}$$

As far as the example of Section 3 is concerned, we first note that the theorem stated above refers to two-parameter families of functions $f_{\alpha,\eta}(x)$, where the parameters are α and η. We interpreted those parameters as agent and environment, but it is clear that nothing depends on this interpretation. What we have in Section 3 is exactly a two-parameter family of functions $f_{b,c}(x)$ (given by (19)) and therefore we can apply the theorem given above and conclude that the graph of the choice function $g(b, c)$ (that is, the set of points (b, c, x) such that x is a global maximum of the function $f_{b,c}(x)$) looks like Figure 13(c) (cf. Figure 11: the Maxwell line starts at the point where $b = c = 1$ and is given by the line of equation $c = 1/b$, for $b \geqslant 1$).

5. Summary

The purpose of this paper was to analyse situations in which similar agents, facing similar or identical environments and having similar objectives, make choices which are considerably different. We first gave two simple examples of this phenomenon, the first where agents face a random variable and have to make a forecast of the value of that variable, the second where workers face a firm with an unknown reservation wage and have to state the wage at which they are willing to work, knowing that they will be employed if and only if the wage they request is below the firm's reservation value. In both examples each agent was assumed to choose the value of a variable x (forecast or wage) so as to maximize (or minimize) his objective function. We showed that agents who were very close to each other (having objective functions which were very close) ended up making very different choices.

In Section 4 we set up a very general framework, in which those examples could be accomodated, and applied to it the classification theorem of catastrophe theory, which essentially says two things: first, that if situations like the ones illustrated in the examples arise, they do so in a stable way (in the sense that the discontinuities involved cannot be eliminated by small changes in the specification of the model); second that (locally) there are only four qualitative types of situation which can arise, namely those illustrated in Figure 13. Moreover this classification into a finite number of qualitative types can be generalized to higher dimensions, for parameter space of up to five dimensions.

The lesson to modellers is as follows. If in an investigation of some problem the data on agents' choices appears to present discontinuities, then the latter should not necessarily be ascribed to 'random noise' and smoothed away by statistical techniques. An alternative approach would be to investigate an appropriate model that would predict such a discontinuity, and then test that model statistically against the data by, for example, applying a least-squares fit to the discontinuous choice graph. Computer programmes for such statistical tests have been designed by Cobb (1978, 1980).

Appendix

Here we give a very simple example of the situation analysed in Section 2, where a unimodal probability function—together with the two-step loss function (4)—gives rise to a bimodal risk function.

Consider the following one-parameter family of skew density functions, illustrated in Figure 14.

$$P_\lambda(y) = \begin{cases} 0 & \text{if } y \leqslant 0, \\ 2y & \text{if } 0 \leqslant y \leqslant 1/2, \\ 2-2y & \text{if } 1/2 \leqslant y \leqslant 1-1/(2\lambda), \\ \dfrac{2(1+\lambda-y)}{2\lambda^2+1} & \text{if } 1-1/(2\lambda) \leqslant y \leqslant 1+\lambda, \\ 0 & \text{if } y \geqslant 1+\lambda, \end{cases} \tag{A.1}$$

where

$$\lambda \geqslant 1. \tag{A.2}$$

It can be seen from Figure 14 that for each λ,

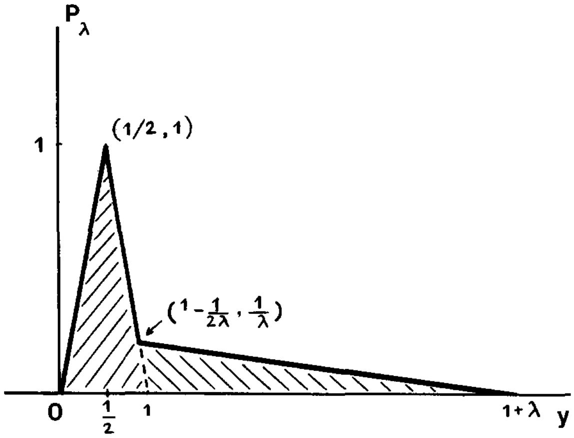


Figure 14. The density function $P_\lambda(y)$

P_λ is the union of two triangles, each of area 1/2 [one with vertices (0, 0), (1/2, 1) and $(1-1/(2\lambda), 1/\lambda)$ and the other with vertices $(1-1/(2\lambda), 1/\lambda)$, (1, 0) and $(1+\lambda, 0)$]. The first triangle represents the most likely value of y (the mode $m = 1/2$), the second represents a possibility of higher y, with likelihood diminishing linearly to zero when $y = 1 + \lambda$. Thus we have

$$\text{mode } m = 1/2, \tag{A.3}$$

$$\text{mean } \mu = \frac{1}{2}\left[\frac{1}{2} + \frac{(1-1/(2\lambda))+1+(1+\lambda)}{3}\right]$$

$$= \frac{2\lambda^2+9\lambda-1}{12\lambda} \simeq \frac{\lambda}{6} + \frac{3}{4}, \tag{A.4}$$

$$\text{ambiguity } \eta = \text{mean sinus mode} = \mu - m$$

$$= \frac{2\lambda^2+3\lambda-1}{12\lambda} \simeq \frac{\lambda}{6} + \frac{1}{4}. \tag{A.5}$$

As in Section 2, we first consider a given density function, that is, we fix a value of λ, say $\lambda = 5$ (in which case, using (A.4) and (A.5) we get $\mu = 1.56$ and $\eta = 1.006$). Let the risk function be given by (6), that is,

$$R_\alpha(x) = \int L_\alpha(x, y) P(y)\, \mathrm{d}y \tag{A.6}$$

where $L_\alpha(x, y)$ is the two-step loss function of Figure 2 and $P(y)$ is given by (A.1) with $\lambda = 5$. As explained in Section 2, we can write

$$R_\alpha(x) = 1 - \alpha B(x) - (1-\alpha)\Gamma(x) \tag{A.7}$$

where

$$B(x) = \int_{x-\beta}^{x+\beta} P(y)\, \mathrm{d}y \tag{A.8}$$

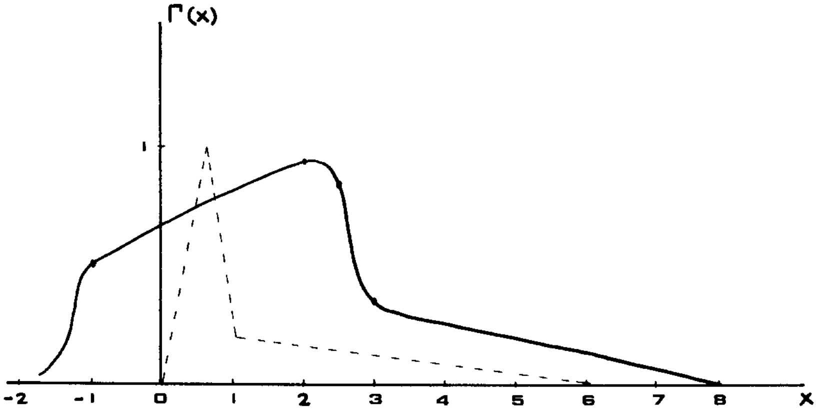


Figure 15. The function $\Gamma(x)$

and

$$\Gamma(x)=\int_{x-\gamma}^{x+\gamma} P(y)\,\mathrm{d}y. \tag{A.9}$$

Choose

$$\beta=0.1 \tag{A.10}$$

and

$$\gamma=2 \tag{A.11}$$

Lemma 1. *$B(x)=0.2\ P(x)$, except within 0.1 of discontinuities of $P'(x)$ at 0, 1/2, 0.9, and 6 where $B(x)$ is smoothed with parabolas.*

Lemma 2. *$\Gamma(x)$ = union of parabolas, as shown in Figure 15.*

If we now let the parameter α vary between 0 and 1, we get a family of C^1 risk functions as shown in Figure 16. It can be seen from Figure 16

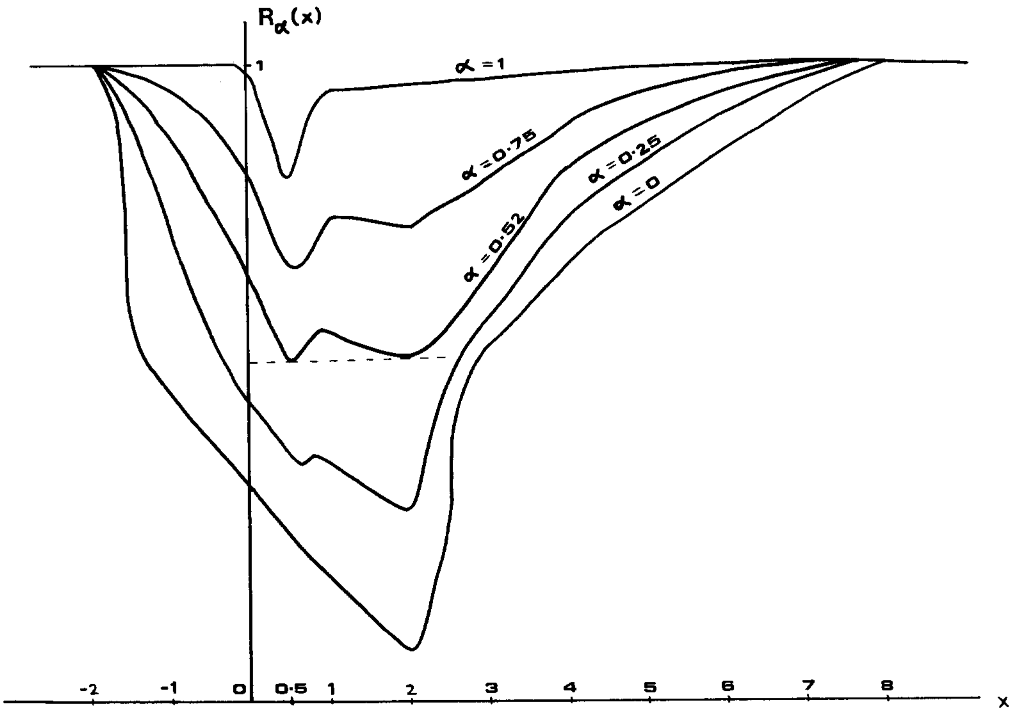


Figure 16. The family of risk functions $R_\alpha(x)$

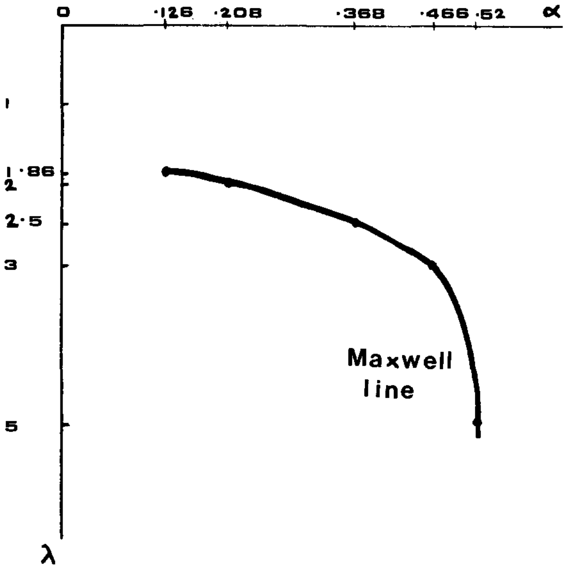


Figure 17. Maxwell line with vertex at $\alpha = 0.126$ and $\lambda = 1.86$

that for values of α close to zero, $R(x)$ is unimodal with a unique minimum near the mean, while for values of α close to 1, $R(x)$ is unimodal with a unique minimum near the mode. For intermediate values of α, $R(x)$ is bimodal with two minima, one near the mode and the other near the mean. The Maxwell point at which the two minima are at the same level is given by $\alpha^* = 0.52$.

If we now let λ vary (subject to $\lambda \geqslant 1$), the Maxwell line, that is, the set of points (α, λ) at which the global minimum of the risk function switches discontinuously from a point near the mean to a point near the mode, is a curve with a vertex at $\alpha = 0.126$ and $\lambda = 1.860$, as shown in Figure 17. Thus we are in the situation illustrated in Figure 13(c).

Note that, although the density and loss functions are piecewise linear, the risk function is smooth (more precisely, differentiable with continuous derivative). The piecewise linearity implies that the risk function does not have second derivative and thus is not generic at the cusp point (the vertex of the curve shown in Figure 17), but arbitrarily small perturbations are.

Finally, as we said in footnote 4, not all skew density functions give rise—together with the two-step loss function (4)—to a bimodal risk function. For example, it can be shown that the following density function

$$P(y) = y/e^y$$

(which is a special case of the gamma distribution) gives rise to a unimodal risk function for every value of the parameters (the same is true for the gamma distribution in general). Thus in this case the Maxwell set is empty and we are in the situation illustrated in Figure 13(a) (continuous choice).

References

Bonanno, G. (1986), "Sudden and surprising changes of attitude during negotiations", paper presented at the First Annual Congress of the European Economic Association, Vienna, August 1986.

Bonanno, G. (1987), "Monopoly equilibria and catastrophe theory", *Australian Economic Papers*, forthcoming.

Chillingworth, D.R.J. (1976), "Structural stability of mathematical models: The role of the catastrophe method", in: J.G. Andrews and R.R. McLone (eds.), *Mathematical Modelling*, Butterworths, London.

Cobb, L. (1978), "Stochastic catastrophic models and multi-modal distributions", *Behavioral Science* 23, 360–374.

Cobb, L. (1980), "Cusp surface estimation", Computer programme, Medical University of South Carolina, Charleston.

Fararo, T.J. (1978), "An introduction to catastrophes", *Behavioral Science* 23, 291–317.

Hirsh, M.W. (1976), *Differential Topology*, Springer, Berlin.

Lippman, S.A., and McCall, J.J. (1982), "The economics of uncertainty: Selected topics and probabilistic methods", in: K. Arrow and M.D. Intrilligator (eds.), *Handbook of Mathematical Economics*, Vol. I North-Holland, Amsterdam, 211–283.

Smith, J.Q. (1978), "Problems in Bayesian statistics related to discontinuous phenomena, catastrophe theory and forecasting", Ph.D. thesis, University of Warwick.

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 by D.H. Fowler, 1975), Benjamin, New York.

Zeeman, E.C. (1977), *Catastrophe Theory: Selected Papers 1972–1977*, Addison-Wesley, Reading, MA.

Zeeman, E.C. (1982), "Decision making and evolution, in: C. Renfrew, M.J. Rowlands and B.A. Segraves-Whallon (eds.), *Theory and explanation in archaelogy*, Academic Press, New York.