Reprinted from the *Journal of Scientific & Industrial Research*, 1977 Vol. 36 No. 3, pp. 103-105

Catastrophe Theory: A New Mathematical Tool for Scientists

MICHAEL BERRY*

Centre for Theoretical Studies, Indian Institute of Science, Bangalore 560012

LET us imagine a landscape whose height is the smooth function $f(x, y)$ of Cartesian ground coordinates x, y, and concentrate on the 'critical points', where f has extrema; these may be maxima (hill tops) or minima (valley bottoms) around which the contours of f are closed, or saddles (cols) at which the contours of f cross (Fig. 1a). Let time elapse. Erosion and deposition will change f and the critical points will move about. Occasionally, the two will collide and annihilate in the sequence shown in Fig. 1 (a–c), or be born and separate (the same sequence in reverse). Such an occurrence is a singularity — it happens only because there is a parameter (time in this case) whose variation changes the form of f.

The general description of singularities of this kind is the subject matter of catastrophe theory, a branch of modern mathematics linking differential geometry. algebra and topology. At its heart is a remarkable theorem developed and proved by the French mathematician Thom[1] and the Russian mathematician Arnol'd[2]. The proof of this theorem is long and difficult. But what the theorem asserts is not hard to understand, and it is also easy to get a feel for th many sorts of situations in which it can be applied. It helps to regard catastrophe theory not as some arcane mathematical medication, a panacea for all problems of science, but as a generalization of the familiar theory of maxima and minima, a useful new acquisition for the physical scientists' toolkit.

Now think of $f(x, y)$ as depending on n parameters, $P_1 \ldots P_n$ instead of just one. Obviously, this introduces richer possibilities for the coalescence of critical points; in other words, singularities more complicated than that of Fig. 1b can occur, resulting from the collision of more than two critical points. The central result of the catastrophe theory is that in a certain sense, to be explained presently, the singularities that can occur for each value of n are finite in number and can be described completely. This 'complete description' is of the whole pattern of development of the singularity as the parameters vary, e.g. the whole of Fig. 1 rather than just Fig. 1b. It is this pattern of development that is called the 'catastrophe'.

Terminology

For each catastrophe, n is the codimension, $P_1 \ldots P_n$ are control parameters defining points in the n-dimensional control space, x, y are the state variables defining points in state space, the form of $f(x, y)$ at the singularity itself is the germ of the catastrophe, the development of the catastrophe out of its germ is the unfolding, and the function $f(x, y; P_1 \ldots P_n)$ is called the potential function. Fig. 1 shows the unfolding of the simplest catastrophe, called the fold[1]. It has codimension 1 (i.e. the control space is a line) and one state variable x (the y coordinate in Fig. 1 is really irrelevant, since all critical points are at $y = 0$). The catastrophe is described by the potential function

$$f(x; P_1) = \frac{x^3}{3} - P_1 x \qquad \ldots(1)$$

with the singularity at $P_1 = 0$ (Fig. 1b), so that the germ is $x^3/3$.

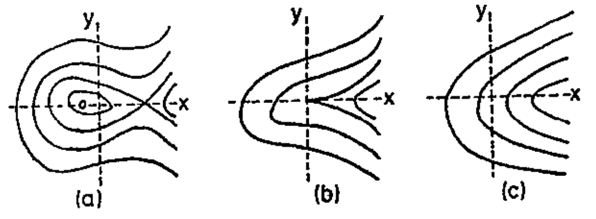

Fig. 1 — The fold catastrophe: Coalescence of two extrema of a function $f(x, y)$ as a parameter (e.g. time) varies

Now follows the promised explanation of the sense in which the catastrophe classification is complete: it describes those singularities that are structurally stable. This means that when the system is subjected to an arbitrary perturbation, the unfolding is topologically unchanged. For example, if the landscape history shown in Fig. 1 had been slightly different, the hills and valleys would be in different places, but the coalescence would still occur (although at a different time) and be describable in terms of Eq. (1) in suitably transformed coordinates. This is a very powerful feature of the theory: it means that the catastrophes are the typical, or the generic singularities that can appear for each n.

Illustrations and Applications

Here are two striking examples of the fold catastrophe in optics: the sun's reflection in water on which there are smooth waves is split into a series of brilliant dancing images. The water surface has the coordinates x, y and the function f whose critical points locate the images is, from Fermat's principle, the optical distance from the sun to the eye via x, y. Now, time is again the single control parameter that changes f as the waves move, so that the images should annihilate and originate in pairs. And so it actually happens, the catastrophic events being called 'twinkles'[3]. The second example is the rainbow, where the state variable x is the point of incidence of one of the parallel sunrays on a raindrop, and the control parameter is the angle θ of emergence of the ray after refraction and reflection in the drop. For given θ, the ray(s) from the drop are determined by the minimization with respect to θ of the optical path function f. For a single internal reflection, there are two rays or none, depending on θ: the two rays coalesce at a particular angle θ_b (which depends on the light's colour) and the bright ring corresponding to this angle θ_b is the rainbow we see in the sky.

This last example shows that it is often the control space that is actually observed, and not the state space as in Fig. 1. For such situations, the catastrophes can be sufficiently described by exhibiting in the control space the regions on which parameters take

---

*Permanent address : H.H. Wills Physics Laboratory, Tyndall Avenue, Bristol B58 ITL, UK.

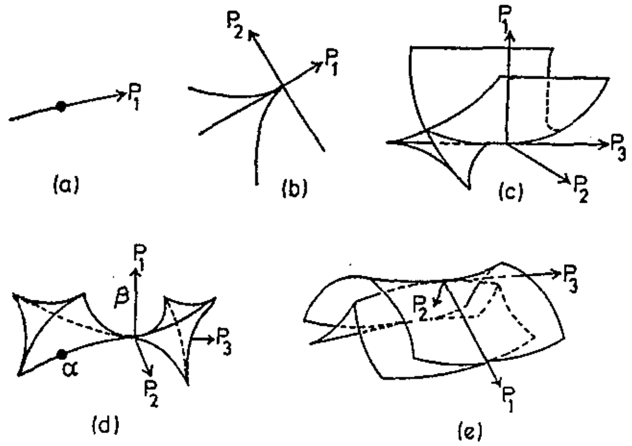

Fig. 2 — Elementary catastrophes in control space for $n \leqslant 3$ [(a) fold, (b) cusp, (c) swallowtail, (d) elliptic umbilic and (e) hyperbolic umbilic]

values causing critical points to coalesce. For the fold this region is trival — a single point (Fig. 2a) on the line P_1. In two dimensions, control space is the plane P_1, P_2 (Fig. 2b) and catastrophes are of two kinds : smooth curves (the trivial extension of the fold into two dimensions) and cusp points. This is far from obvious, and reveals that singularities in P_1, P_2 in the form of isolated points, or corners with a finite angle, or lines coming to an end, are nongeneric — they are structurally unstable and the smallest perturbation will break them up into the generic fold lines and cusp points. For codimension three, the catastrophes can be fold surfaces or cusp lines in the control space P_1, P_2, P_3, and also singular points of three different types : the swallowtail (Fig. 2c), the elliptic umbilic (Fig. 2d) and the hyperbolic umbilic (Fig. 2e). For $n > 3$, there are more complicated shapes[1,2].

These rather unlikely forms find immediate application in the classification of caustics. These are generalized focal regions, the envelopes on which neighbouring rays touch. They are catastrophes, because the rays are critical points of an optical path function and their touching corresponds to a coalescence of critical points; control space is the actual space in which the rays travel. The rainbow has been already mentioned as a fold caustic. The cusp is a familiar sight in any brightly illuminated teacup or shiny concave vessel. A highly non-trivial example again concerns sunlight and water waves, but this time not the reflections, rather the moving pattern of caustic lines formed by refraction on the bottom beneath the waves, if the water is not too deep. These lines often meet[4] in what appear to be triple junctions (Fig. 3a). This would present something of a problem for catastrophe theory, since one can hardly imagine a more generic function f than that generated by random water waves : yet a triple junction of lines is nowhere to be found among the generic catastrophes of Fig. 2. The triple junction apparently observed might really be a much finer generic structure, imperfectly resolved; in an elegant experiment (unpublished), Prof. N. F. Nye showed that under high magnification the junctions are indeed generic in form; Fig. 3b shows a typical example of one of the caustics observed by him.

In optics, the catastrophe theory goes far deeper than this : as well as classifying the caustics of ray families, it classifies the diffraction patterns that soften the singularities of geometrical optics when the wavelength is small, but not negligible[5,6], marking a major step towards the solution of the old problem of the short-wave asymptotic form of solutions of Maxwell's equations. Moreover, any type of wave can be studied in this way; in particular, the details of the classical limit of quantum mechanics hatve been clarified. Fig. 4 shows the contours of the intensity of the diffraction function[7] corresponding to the cusp catastrophe.

Here now is an application of the theory in which the state and control spaces are equally important. Let the contours of $f(x, y)$ (Fig. 1) represent streamlines of a flowing liquid with a dissolved polymer. Catastrophes are produced by altering the flow pattern, so that vortices and stagnation points (saddles) are born or coalesce. The stagnation points, where two streamlines enter and two leave, are of particular interest as sources of high extension of polymer chains dissolved in the liquid[8]. Fig. 5a shows a higher order stagnation point with three ingoing and three outgoing streamlines; this flow can be produced in a 'six-roll mill', i.e. between six

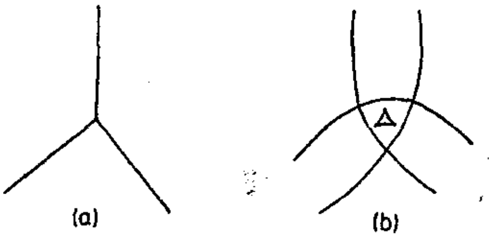

Fig. 3 — (a) Apparent triple junction (nongeneric) frequently observed in caustics of refraction from wavy water surfaces, (b) central region of (a) under high magnification as observed by Prof. J.F. Nye, for typical case

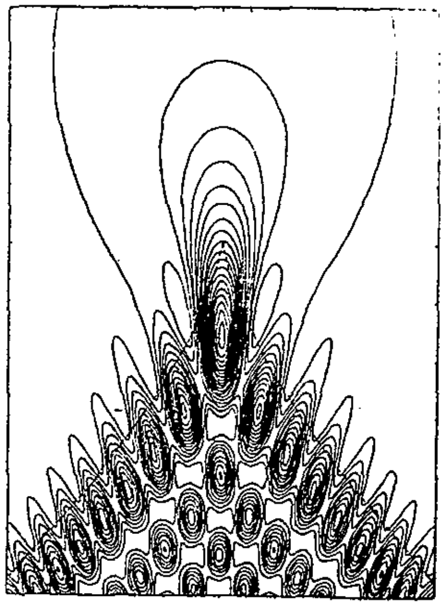

Fig. 4 — Contours of intensity of Pearcey's function,[7] for diffraction near a cusp

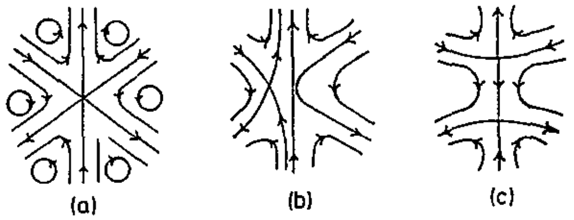

Fig. 5 — (a) Symmetrical singular flow in six-roll mill, corresponding to the origin in Fig. 2d, (b) singular flow in six-roll mill, corresponding to α on Fig. 2d, (c) nonsingular flow in six-roll mill, corresponding to β in Fig. 2d

cylinders rotating in the senses shown. It is obviously unstable to perturbations, such as differences in roller speeds. Now, the stream function f, whose contours are as in Fig. 5a, is the germ of the elliptic umbilic potential function, whose full unfolding is[1]

$$\frac{x^3}{3} - xy^2 + \frac{P_1}{2}(x^2 + y^2) + P_2\, y - P_1\, x \qquad \ldots(2)$$

Since this function is structurally stable, it follows that all possible flows close to that of Fig. 5a can be generated with just three control parameters, P_1, P_2 and P_3. These can be directly related[9] to the roller speeds, so that only three need be independently controlled. By changing these speeds, a path in the control space of Fig. 2d can be described. Every time the elliptic umbilic surface is crossed, the corresponding flow is singular, i.e. at least two of its critical points coalesce. In beautiful experiments, Mackley[9] produced all topologically different flow patterns contained in Eq. (2). Fig. 5 shows three examples : 5a corresponds to the control point at the origin of Fig. 2d, 5b corresponds to the control point α on the cusp edge in Fig. 2d, and 5c corresponds to the non-catastrophic control point β. These experiments did not merely illustrate the kinematics of the flow (2), but also gave valuable and surprising[9] information about the rheology of polymer solutions.

Structural engineering has its share of catastrophes in the everyday sense of the word, and it has recently become clear that where these involve sudden buckling, they are also catastrophes in the mathematical sense. The state variables x, y are co-ordinates of struts, beams, etc. and the control parameters $P_1 \ldots P_n$ are the magnitudes and positions of the applied loads. The potential function $f\,(x,\ y;\ P_1 \ldots P_n)$ is the actual potential energy of the system. When minimized with respect to x, y, this gives the state of the system for fixed controls. As the controls change, the state changes smoothly, except at control points for which extrema of f coalesce; then a minimum disappears and the system jumps catastrophically to a different minimum. The catastrophes thus describe all the generic modes of failure of such systems. This is of obvious practical importance, particularly in view of the fact that the forms of the higher catastrophes are far from obvious. Complicated engineering problems have been examined in this way[10]. It must be stated, howver, that to date no results have been obtained that were not already known from the more laborious and less elegant techniques of bifurcation theory. Therefore, these studies can be classified as illustrations rather than applications of the catastrophe theory. The same remark applies to the analysis of phase transitions using the catastrophe theory : it is known[11] that all forms of Landau mean field theory can be regarded as catastrophes, but physicists are unlikely to be convinced of the usefulness of this new mathematics by what they see as a highly abstruse derivation of van der Waals equation[12].

Here, the potentially far-reaching applications of the catastrophe theory in the 'soft sciences' — biology, sociology, economics, linguistics, etc., have not been mentioned. The case for these is argued passionately and poetically in the beautiful but infuriatingly obscure book of Thom[1], and these topics have also recently been well reviewed in a more down-to-earth manner by Zeeman[12].

For physical scientists, the novelty and promise of the catastrophe theory lie in the concepts of genericity and structural stability. These make it possible to discuss the details of symmetry-breaking — at least for those systems (involving minimization of a smooth function) to which the mathematics applies — since by the application of special methods it is often possible to solve a symmetrical (structurally unstable) particular case of a problem and identify the solution as a section through a catastrophe, whose general form then gives the (structurally stable) solution when the symmetry is absent[13]. This procedure fails for high orders of symmetry where many control parameters are required for the generic unfolding; this is because catastrophes have not yet been classified for arbitrarily large codimension n. In particular, when the symmetry is continuous (e.g. a rotation), the codimension is infinite and the 'generification' of such cases suggests new directions for mathematical research[5].

The last observation highlights an interesting complementarity that often exists between artefacts of man, deliberately constructed with high symmetry (optical instruments with an axis of rotation, cylindrical shafts, etc.), and natural objects, structurally stable and having only topological similarity and no high symmetries (irregular refracting water surfaces, trees, skeletons, etc.).

Acknowledgement

I thank Prof. K.P. Sinha for the hospitality of the Indian Institute of Science where this review was written.

References

1. THOM, R., *Structural stability and morphogenesis* (Benjamin. Inc., New York), 1975.
2. ARNOL'D, V.I., *Funct. Anal. Appl.*, 6 (1972); 3-25; 7 (1973), 75-6.
3. LONGUET-HIGGINS, M.S., *J. opt. Soc. Am.*, **50** (1960) 838-56.
4. MINNAERT, M., *The nature of light and colour in the opne air* (Dover Publishing Corp.), 1954, Plate IV.
5. BERRY, M.V., *Adv. Phys.*, **25** (1976), 1-26.
6. DUISTERMAAT, J.J., *Communs. Pure appl. Math.*, **27** (1974), 207-81.
7. PEARCEY, T., *Phil. Mag.*, **31** (1946), 311-7.