TIM POSTON

On deducing the presence of catastrophes

Mathématiques et sciences humaines, tome 64 (1978), p. 71-99

<http://www.numdam.org/item?id=MSH_1978__64__71_0>

Math. Sci. hum. (16e année, n°64, *1978,* p.71-99)

ON DEDUCING THE PRESENCE OF CATASTROPHES*

Tim Poston**

"With their continuity (personality) fallen into the false view of two extremes, all the heretics, who, because they have not understood their own discriminations, are people without spiritual success, have constructed talk about impermance."

Lankavatāra Sūtra (trans. E. Conze)

CONTENTS

* This work was supported in part by the Fonds national suisse de la recherche scientifique grant No. 2.461.0.75

** Battelle-Genève

1. INTRODUCTION

The purpose of this note is discussion in somewhat more detail of an argument used by Zeeman, in [33] and elsewhere. Zeeman's paper being directed primarily at a non-mathematical audience, the topological niceties of the argument were quite properly omitted: and since the possibility of a detailed completion of the argument is apparent to any experienced differential topologist, it has not seemed necessary until now for anyone to elaborate it in writing. However, there has now appeared a claim, by authors with credentials as mathematicians, that no such argument is possible.

Much of the discussion in the papers [25, 26, 27, 31] of Sussmann and Zahler is concerned in fact, though not in appearance, with general questions of mathematical modelling rather than of catastrophe theory; for example, whether it is legitimate in principle to describe with a continuous variable an entity known to be fundamentally discrete. (If it is not, most of population dynamics, nuclear reaction rate theory etc. are illegitimate along with some catastrophe theory.) To some extent their case may be permitted to fall of its own weight: representative arguments being appeal to the character of attack motivated by hunger ("a tiger slowly jumping at a deer" [26§4a]) in discussing Zeeman's model of behaviour where Lorentz [13] had gone to great lengths to make rage and fear outweigh all other factors; bald assertions*;

---

*) Such as "The reader should try to imagine a continuous progression from snarling to attacking... He will fail. No matter how good his imagination is, the progression will be discontinuous: there will be one instant where the dog jumps." [26§4a] (where exactly, in the easily imagined progression snarling → circling close → snapping at the heels → nipping at the heels → biting at the ankles → biting hard at the upper legs → rearing up to bite the arms → going for the jugular?); "a dog A will not jump at ....

and quotation from inaccurate brochures advertising encyclopaedias [27].. (They do in [31] quote Stewart's Britannica article [23] itself - "Properly understood and exploited, this ever-expanding web of concepts promises.... a profound insight into the universe" - then say "We disagree" on the basis of [26] recapitulated plus a detailed attack on a paper [10] in which they maintain catastrophe theory is not properly understood since an inadmissible surface is drawn, though Fig. 2 of [10] and its description should be compared with their redrawing of it. Their disagreement would be better based on the physical examples adduced as solid by Stewart than on those which an encyclopaedic overview must describe, but for which he remarks [23, 3rd p.] that "much criticism regarding their utility has yet to be assessed".)

However when professional mathematicians claim, in a long paper in a learned journal, that a major and systematic mathematical error has been committed, more detailed discussion is needed than for their response with questionable assertions to some tentative models. I do not wish to argue the positive truth of all Zeeman's more speculative models, some of which I find almost as unconvincing as the counter-models in [26] and some of which are certainly

---

.... another dog B, and sink its teeth one millimetre deep into B's skin, having achieved nothing but getting B angry" [26§5] (the only kind of dogfight I have ever witnessed, fights to the death being uncommon); and in [26§13] "History presents us with many cases of administrations that stubbornly refused to end a war despite the fact that the majority of the population wanted it ended... but few cases of governments that refused to go to war despite strong pressure by public opinion." What book lists wars that did not happen? Certainly peoples have pushed unwilling governments to war. Pitt, hearing the popular rejoicing at achieving war with the Dutch, said "They are ringing the bells now: soon they will be wringing their hands" - but if he had prevented the war no such memorable story would exist. Popular 'By Jingo' feeling created in some measure the Crimean war, which the government then conducted with such awesome incompetence as to suggest to an analyst that "one is forced to consider the hypothesis that at some level in the minds of those who direct national aggression there lurks a contrary motive, a need to pull their punches." [5p.66] intended mainly to convey what kind of behaviour a cusp catastrophe, if present, can imply. Like Sussmann [25] I am a "professional sceptic" about mathematics in the behavioural sciences, and take their methodological difficulties more seriously than I perceive either Sussmann or Zeeman as doing. (For some exploration of these difficulties in relation to catastrophe theory, see [19].) I am chiefly concerned in this paper with the legitimacy of mathematical deductions from hypotheses once (by whatever means) these are chosen, a necessary step in most sciences towards testing the hypotheses themselves.

The point at issue is the following. Sussmann and Zahler assert [26§7] that it is false that "by means of Thom's theorem, one can draw conclusions about the critical set even for those values of [the control parameters] for which the hypotheses specify nothing"; conclusions which Zeeman certainly claims to draw. Similarly in [25] Sussmann, discussing behaviour over a range of a parameter C, remarks correctly that Zeeman's hypotheses concern only values $C < C_1$ or $C > C_2$ for some $C_1 < C_2$. Then "Zeeman is assuming nothing about what happens for $C_1 < C < C_2$, and claims to <u>deduce</u> the behaviour for C in this middle range. Such a deduction is, however, impossible, since any behaviour whatever... is compatible with Zeeman's assumptions for $C < C_1$ and $C > C_2$". Similar counter-assertions may be found at various other points in [26].

Now, given the general setting of differentiable functions*, these statements are already remarkable. Few mathematicians, informed analogously that a smooth function f had the value -1

---

*) Sussmann and Zahler also object to the use of continuous variables and differentiable relationships, but these objections are (i) explicitly distinct from their attack on Zeeman's deductions <u>within</u> this setting (ii) a reality-modelling, not a mathematical, question.

at $C_1 = 1$, the value $+5$ at $C_2 = 2$, would hesitate to assert the existence of points $C_1 < C < C_2$ where f takes the value 0 or where df/dC takes the value 6 . (Few scientists looking at particular systems would fail to adopt as a mild additional workin hypothesis that such points are unique and distinct.) Zeeman is using deeper mathematics than the Intermediate or Mean Value Theorem, but this example, like the sophisticated use of fixed point theorems in functional analysis, serves to illustrate that topological argument from data <u>at</u> points to behaviour <u>between</u> points is a cornerstone of modern mathematics, pure and applied. Certainly there are instances where continuous systems beget discontinuity - indeed, an interesting case where Courant and Robbins [4] appeal inappropriately to the Intermediate Value Theorem is analysed in [16] - and the whole point of catastrophe theory is to analyse this phenomenon for a wide class of systems. But this still leaves some surprise at the proposition apparently being made here, that topological deduction about behaviour in the middl from information about the ends is inherently impossible.

However, since the completion of Zeeman's argument is of distinctly greater depth than an appeal to the Intermediate Value Theorem - it does in fact require the full strength of the cluster of results known as Thom's theorem - it remains necessary to establish that this particular form of argument 'inwards' is indeed possible. To do so I will take the instance in which Zeeman's mathematical hypotheses are most explicit, and to which Sussmann and Zahler in [26] devote most space: the discussion of frontier formation in [33]. But before considering this in detail since the argument will require repeated appeal to the genericity of transversality, it is necessary to separate this principle from the misrepresentation of its character in [25] and [26].

2. GENERICITY

It is certainly true that one aspect of genericity for a property $\mathcal{T}$ in a space X is density (that every P in X can be arbitrarily closely approximated by P' with property $\mathcal{T}$) and that this "approximation principle" is the aspect Isnard and Zeeman [9] chose to explain in a footnote. It is also true, as observed in [26§12], that density of $\mathcal{T}$ is not of itself a justification for assuming $\mathcal{T}$ effectively true of all relevant Ps. One relevant example is that the set of all irrational numbers between 0 and 1 is dense - indeed, the countable intersection of open dense sets, i.e. generic, and moreover of measure 1 - but it would be absurd to argue that the qualitative properties of irrationals are in some useful sense universal. (Furthermore, within catastrophe theory one rapidly learns that often a designed 'ideal' system is not generic*. Since it can then be approximated by many, qualitatively different, generic systems the appropriate response is not to use any one as an approximant - as [25§7a] would suggest that catastrophe theory

---

*) The abus de langage principle that "generic f" as distinct from generic property, "is usually taken to mean that f has a lot of generic properties" [21] appears in the paper that first suggested 'generic' apply strictly only to properties, and has remained standard in the technical literature (see, e.g. [2]). The cavils in [26§12] at its use, in papers which try not to dazzle the non-mathematician with too many technical definitions, are thus puzzling. A good case could be made for calling generic an object possessing all natural generic properties, where these could be proved countably many: indeed, one might require generic properties to be functorially natural, to exclude frivolous examples like 'not getting up at 8 o'clock' [26§12], and the fact that for any sets $Y \subseteq X$ a topology can trivially be tailored to make membership of Y generic. In most important cases naturality is clear by inspection, as it is for Sussmann's proof [24] that certain systems are generically 'controllable'. (It is fair to ask why this result was published: if "Thom's theorem is inherently uninformative" [26§7a], so by the same reasoning is Sussmann's.)

does - but to unfold it universally and so consider all possible nearby systems. For an example where the traditional, unstable, non-generic one-parameter functional analytic approach to compound plate buckling is replaced by an 8-parameter structurally stable treatment in this way, with new behavioural predictions, see [14]. For quantitatively successful examples in scattering theory, fluid mechanics, laser optics etc. see [19].) Genericity has thus often seemed an unimpressive property to - for instance - physicists hearing from topologists about new results in dynamical systems theory or its subset, catastrophe theory. Many topologists have, however, given immense attention to it: not out of failure to perceive that genericity need not be informative, but from a clear perception of why in the present context it is.

Most genericity statements by topologists are closely related to, or derivable from, the Thom transversality lemma ([1], [11] detail this relationship): the fact that in various circumstances transversality is generic. It is a pity that, because genericity is a formally provable expression of the way transversality is typical, and its failure for some circumstances (symmetry, modality) an excellent technical test for transversality's ceasing to be typical, genericity has generally been chosen as the means of communication of this typicality to non-mathematicians. (One is reminded of the French schoolchildren introduced to the 'angle' concept by way of the inverse sine function, for Bourbachique technical convenience.) It is a pity because not only have physicists already believed transversality typical, without topologically expressing the fact, for generations: they have been extracting quantitative consequences from it with spectacular regularity.

To prove the first of these assertions one could quote from countless scientists: choosing Eddington as a representative, for his usual clarity of expression, we find in a 1918 lecture [6] "In two dimensions any two lines are almost bound to meet sooner or later; but in three dimensions, and still more in four dimensions, they can and usually do miss one another altogether, and the observation that they do meet is a genuine addition to knowledge." This idea, quite exactly, underlies the typicality of 'nothing but fold catastrophes' for a 1-parameter family of functions: for a detailed tracing of the connection in informal geometric language see [18] or [19]. (The rest of Thom's theorem relies on precisely the same idea in higher dimensions.) This connection thus translates Eddington's point about "a genuine addition to knowledge" to Thom's statement [29] on catastrophe theory's predictions that "If the prediction is realised, there is nothing to be surprised about. If the prediction fails (that may happen) and a morphology M_{12} different from M_1^2 does appear, this is interesting, because it shows that our original assumptions were too simple, and some new element of complication has to be introduced into the picture." The principle, not localisable in either science or mathematics but rather an aspect of the strange resonance between the two, is the same for Thom as for Eddington: the difference is only that Thom has technical tools by which it may be pursued to draw conclusions beyond those accessible to Eddington. Both quotations above treat it as susceptible to experiment, and in the same way. Both are far from "claiming that the world can be deduced by pure thought", as [26§7a] describes the transversality principle.

In [25§5] Sussmann cites an excellent warning by Thom [28] against overconfident numerical modelling with catastrophes (which, given current practice in the soft sciences, could usefully be extended to models with normal distributions, linear correlations, or control theory)

> "Many people, understandably eager to find for Catastrophe theory an experimental confirmation (?), may embark into

precarious quantitative modelling, where explicit observable interpretations are given to unfolding parameters (even to internal ones..). Needless to say, many (if not all) of these interpretations will break down. This may cause - among positivist-minded Scientists - a "backlash" reaction against Catastrophe theory, a reaction already noticeable among some scientists in UK. (In France and in the US, Catastrophe theory is still too ignored to have provoked such a reaction.) In the same line, I would also like to add a didactic warning: when presenting CT to people, one should never state that, due to such and such a theorem, such and such a morphology is going unavoidably to appear."

omitting the three succeeding sentences, in which Thom (after a remark about all mathematics in all sciences) restates the very well tested principle we have been discussing:

"In no case has mathematics any right to dictate to reality. The only thing one might say is that, due to such and such a theorem, <u>one has to expect</u> that the empirical morphology will take such and such a form. If reality does not obey the theorem - that may happen - this proves that some unexpected constraints cause some lack of transversality, which makes the situation all the more interesting."

Sussmann claims the part that he does quote as a "second proof" that no deductions from the principle can be made about behaviour in the middle range, and after a paragraph on genericity as an approximation principle says "... hence CT is not testable. Thom is clearly aware of this, as shown by his statement quoted above, but Zeeman is not". Catastrophe theory is testable, and Thom's interest is explicit in the unexpected constraints that can be revealed by such tests. (Compare the physicist's interest in the discovery of 'superselection rules', revealing that more symmetry constraints had to be admitted to particle theory than had been supposed.)

To illustrate the point that the transversality principle, unformalised, is a longstanding quantitative tool of physics, let us consider the following argument from Landau and Lifshitz [12 p. 436]:

> "If B does not vanish identically [for symmetry reasons] then the transition points are determined by the two equations: $A(P,T) = 0$, $B(P,T) = 0$. Hence in this case the points of continuous phase transition are isolated ones."

This "hence" is precisely an appeal to transversality, since if the mapping $(A,B) : R^2 \rightarrow R^2 : (P,T) \mapsto (A(P,T),B(P,T))$ is not transverse to $\{(0,0)\}$ the common zeros of A and B can be any closed plane set whatever, from a logarithmic spiral to a portrait of the physicist as a young man. But, <u>given transversality</u>, they are isolated. The sense in which Thom's theorem says "almost all" is exactly that in which much physical writing says "all" for the same reasons. I have sometimes been made uneasy by Zeeman's use of "all" in this context, feeling it could be misleading: Sussmann and Zahler show that it can be equally misleading in nontechnical exposition to attempt greater precision, since they achieve a deeper misunderstanding of the qualifier "almost" than I have yet seen of the unqualified "all". On a tightrope between pure mathematics and sociology, the level of rigour of a standard physics text is perhaps a reasonable compromise.

Landau and Lifshitz go on to show, using heuristic Taylor expansion truncations where a topologist would now use determinacy and versality criteria, that the isolated points above are specifically what a catastrophe theorist would call cusp points. They then draw a number of quantitative conclusions from this, notably in the form of 'critical exponents'. (Compare the relation "slope of recession = $L(C-C_0)^{\frac{1}{2}}$ where L is a non-vanishing smooth function of C" [34] about which Sussmann [25§5] says "Zeeman's error [of deducing a cusp present at all] is compounded" by a claim that the cusp has the standard algebraic form in the original coordinates. No such claim appears in [34] - the canonical formula is quoted with reference to a figure, which must needs be described by some coordinates - and the relation cited does not require it, the asymptotic exponent ½ being an invariant of the cusp under the smooth transformations admitted. Zeeman's explanation of this point is described in [25§5] as "he appears to withdraw the claim" that he does not make: though he may seem to make it when the above quotation from [34] is given, as Sussmann does, without the words I have underlined.) These exponents are not only testable: they caused a major crisis in the physics of condensed matter when they turned out, experimentally, to be wrong!

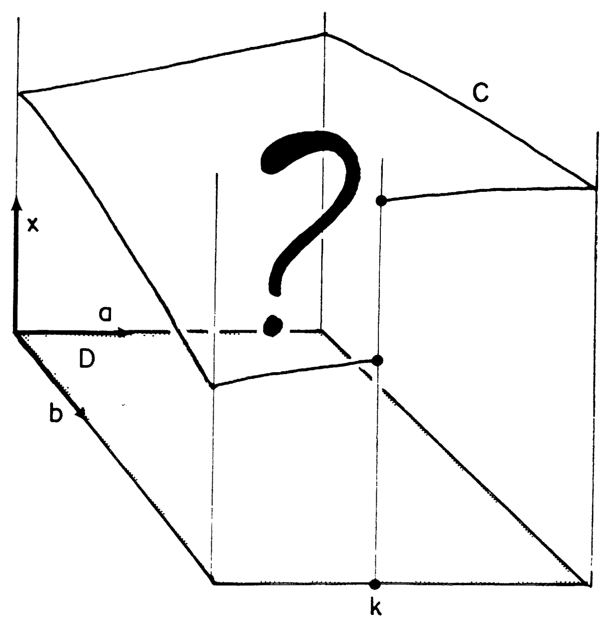
The error, however, turned out to be not in the appeal to the transversality principle but in the approximations assumed for the behaviour of matter around a degenerate minimum like $(x^4+O(5))$ for a thermodynamic potential. The fluctuations at such points are better treated by the new 'Renormalisation Group' methods. These apply with experimentally good results to obtaining information both from 'exact' statistical models and from smoothed heuristic ones, in whose choice the transversality principle - unformalised - plays the same rôle as in the Landau theory. Moreover the Landau theory remains an important tool in giving technologists numbers, fairly reliably away from the close vicinity of the transition point, and for much less computational effort than the more refined Renormalisation Group approach.

Note the warning of this case for overeager naive application of catastrophes: getting right the differential topology which is the 'pure mathematical' end of catastrophe theory, even where catastrophe theory is importantly applicable, does not guarantee correct predictions unless the analysis of how the catastrophe governs the matter (fluid, economy, foetus...) concerned is a good theory in the scientific sense. (The relationship is equally subtle for scattering problems: see [3] or [19] for accounts.) Both ends matter; where phase transition involves higher singularities (as at 'tricritical points') one must substitute the formal tools of catastrophe theory for physical rule of thumb to get mathematically correct results [19].

3. THE EXISTENCE OF A CUSP

Consider now the details of deducing, for points where it is not specified in advance, the singularity structure of the set of equilibria. The general mathematical hypotheses are the transversality principle, and that the equilibria are invariably points (not, e.g., limit cycles). This latter is not an automatic consequence of transversality except in the setting of gradient dynamics, but it is true often enough to be a common working hypothesis in scientific theories, and is far more general than the gradient assumption: one need not assume that the dynamics is driven by minimisation. (Technical aside: the elementary catastrophes can govern even <u>non</u>-point-attractor bifurcations, as in [7]; but language is complicated here by such examples as the Hopf bifurcation from 'gradient-like' to 'non-gradient-like' dynamics, which can stably be reduced to the bifurcation of a Liapunov pseudo-gradient function but not to an ordinary elementary catastrophe of finite codimension. Thus there can be no "appeal to Thom's theorem" as such where there are limit cycles, though the stable phenomena it classifies remain important.) The bifurcation of such attractors can commonly be shown equivalent to those of minimisation problems, even if that of their basins of attraction cannot. The choice of minimised function is not unique, but the implications of the transversality principle correspond in the two settings. Hence I shall argue in the case where 'equilibrium' means 'critical point of a real-valued function V', bearing in mind that V may be a construct to help analyse a bifurcating differential equation, rather than a quantity with quasi-independent physical 'meaning' (like wave phase, free energy, or utility) for which applicability of Thom's theorem is more obvious.

We are concerned to show that if for control values (a,b) on the boundary ∂D of the rectangle D in Fig. 1 there is a continuum C of stable equilibria with a discontinuity at k in the state variable $x \in X$ (X is drawn 1-dimensional for graphical reasons), then there is a cusp point for some control value (a,b) <u>inside</u> D . (For definitions of 'catastrophe theoretic' terms see [35], [19] or even [26].) Let us denote the equilibrium set, for all $(a,b) \in D$, by S and its boundary, containing C , by ∂S . It is part of Thom's theorem, or easily derived directly from the transversality principle, that S is a surface, ∂S a curve. (In [26 § 7a] we are told Thom's theorem "is not much better than the tautologous statement that an equilibrium surface is a surface",



<u>Fig. 1</u>

but if transversality is "inherently uninformative" there are no grounds for supposing there is a surface to make a statement about. "The equilibrium set is a surface" is importantly nontautologous and fails when transversality does, e.g. if symmetries are present: for many examples see [19].) I assume that for any (a,b) there are finitely many equilibria and that no equilibrium 'goes off to infinity': with Thom's theorem forbidding local pathologies, this means that I suppose S a compact surface. The 'catastrophe map' χ_V that projects S to D must certainly have some singularities: if S had only nondegenerate equilibria the local description of it via the Implicit Function Theorem* as the graph of a function $x = f(a,b)$ could be pieced together (D being simply connected) to describe it globally as the graph(s) of one or more smooth functions, contradicting the discontinuity at k. What must the singularities be?

In some sense, the presence of a cusp is obvious. [Indeed, it might be argued that footnote 7 of [26] accepts - not merely attributes to Zeeman - that "it follows that there must be at least one cusp". But then either one can draw conclusions as denied elsewhere in [26§7] (see Introduction above) or "there must be at least one cusp" is not a conclusion although "it follows".] But the Jordan Curve Theorem, for instance, is rather more "obvious" and is not only beyond most undergraduates to prove but has an equally "obvious" generalisation to three dimensions which - to general surprise - turned out false. The treatment of the presence of a cusp that follows is as simple as I can devise, but it does need the full strength of the cluster of results (on

---

*) This result is used in [25] and [26] without explanation or reference, apparently assumed not "mystificatory" ([26§8]), though known by name in my experience to rather few scientists, even among the more mathematical groups such as theoretical engineers. An informal account may be found in [19], a proof in e.g., [20].

transversality, reduction of Taylor series to polynomial form, enumeration of possibilities, etc.) that make up the omnibus entity "Thom's genericity, structural stability and classification theorem".

First I must make explicit one more boundary assumption, T : for no $(a,b) \in \partial D$ are there more than three equilibria. This rules out S as in Fig. 2, which is generic and in which no cusp point appears. (Assumption T is clearly made in some of the papers where Zeeman deduces a cusp, notably in discussions of 'bi-modality': in others Fig. 2 could be excluded by invoking other hypotheses he states. In any case my concern here is less with particular models - to some of which I have independent objections - than with the general allegation in [26§7a] that "The claim that Thom's theorem can be used to infer the shape of a surface from some partial knowledge of it is false". It would be interesting to know, if Sussmann and Zahler as quoted in the previous paragraph were indeed agreeing that "it follows", by what specific hypotheses they agreed to rule out Fig. 2.)

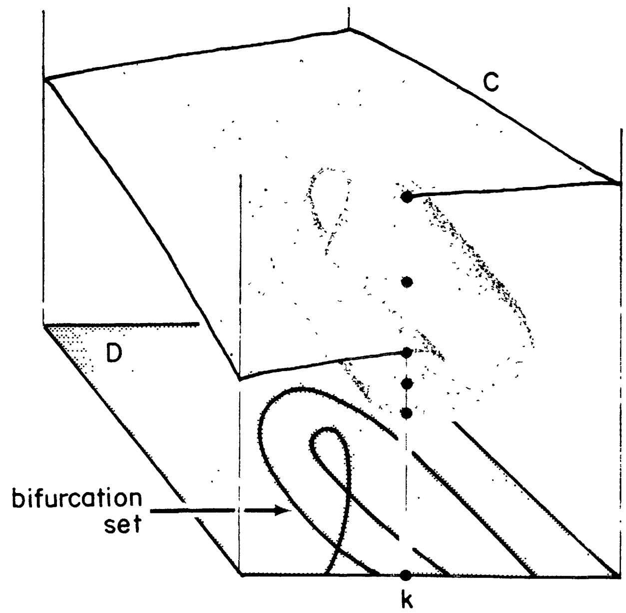
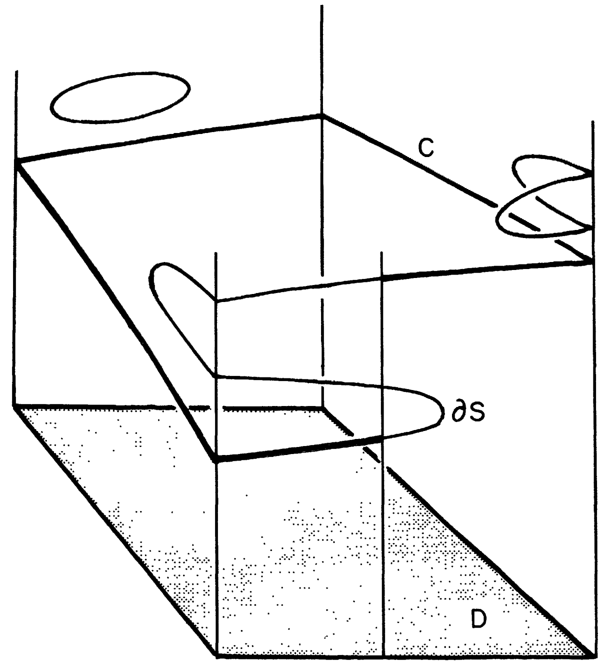


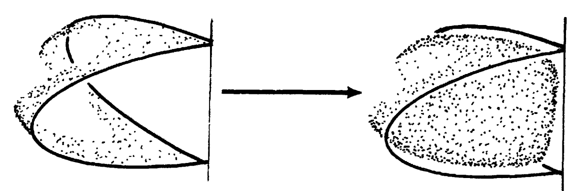
Fig. 2

Now, suppose that we can assume without loss of generality that X is one-dimensional. (Sussmann and Zahler describe this in [26§12a] as "probably the only step in the entire "proof" that can be given a sound justification", but see below.) By the transversality principle via Thom's theorem (from now on I will shorten this to 'by Thom's theorem'), ∂S has only regular and fold points.

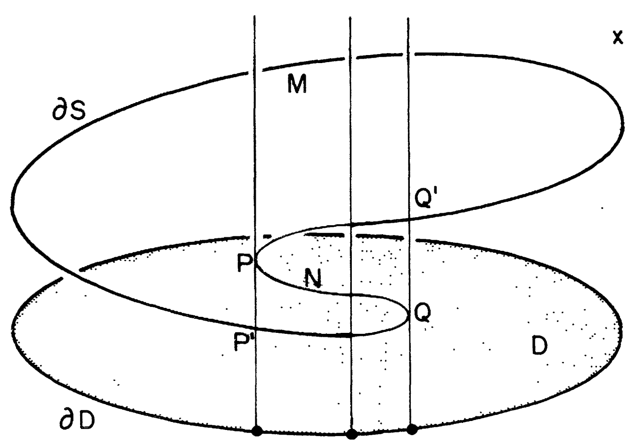
If it has <u>no</u> fold points, assumption T and the jump at k imply that ∂S is either a loop (containing C) wound twice onto ∂D , perhaps with another, single, loop, or a loop wrapped three times around. (This last would require that for each $(a,b) \in \partial D$ there are three stable equilibria and no others. A little surprisingly, this is possible; $V(x,y)$ defined on the plane can have exactly three critical points, all non-degenerate minima.) These configurations are realisable for a control space like ∂D , topologically a circle, but not under our present hypothesis that X may be assumed one-dimensional, which would force ∂S to cross itself: the reasoning leading to Thom's theorem includes the information that the catastrophe manifold sits above the control space without self-intersections. (Nor are they realisable ever as ∂S for S over simply connected D unless S has cusp as well as fold points, but I shall not prove this here.) It follows easily that ∂S must fold back on itself as in Fig. 3, perhaps with disconnected loops as shown. These may



<u>Fig. 3</u>



<u>Fig. 4</u>



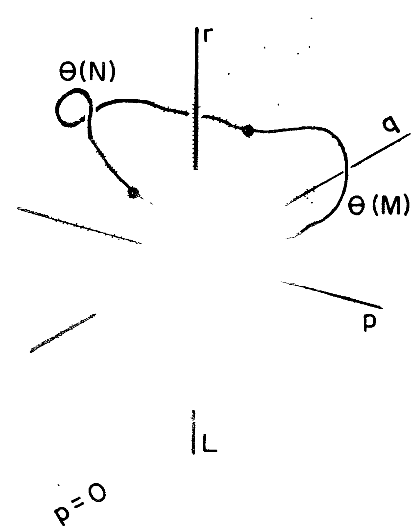
<u>Fig. 5</u>

be 'pinched off' as in Fig. 4 (since the universality of the unfoldings given by ∂D as control implies that 'collars' as in Fig. 4a must always exist) without introducing a cusp. Thus without loss of generality I may argue in the case shown in Fig. 5 (where D is taken as a disc to avoid irrelevancies with the positions of p and q relative to the corners, which may be removed by obvious smoothing arguments.)

There are four natural 'pieces' now to ∂S : the curve M of minima, the curve N of maxima folded between them, and the two fold points P , Q . Now, if we do not permit orientation-reversing changes of variable, we see that V at P must be equivalent to $+x^3$, since in approaching the minimum at P' from P both x and V decrease, and V at Q must be equivalent to $-x^3$ since V decreases to $V(Q')$ by <u>increasing</u> x ; we know the general form x^3 by Thom's theorem. (The signs would reverse if x increased 'downward' in the figure, but would remain opposite.) Then if at each point $s = (a,b,x)$ of ∂S we Taylor expand $V_{a,b}$ around x to order 4 , we get a polynomial

$$\text{constant}_s + 0x + p_s x^2 + q_s x^3 + r_s x^4$$

with coefficients depending on s , the linear term vanishing by the equilibrium condition. Discarding the s-dependent constant, we define a map $\theta : s \longrightarrow (p_s, q_s, r_s)$ from ∂S into the space I_1^4 of polynomials $px^2 + qx^3 + rx^4$. Then M is mapped to points with $p > 0$, N to $p < 0$ and P , Q to points $\theta(P)$, $\theta(Q)$ on the plane $p = 0$ where q has opposite signs (since one reduces to



<u>Fig. 6</u>

x^3 , the other to $-x^3$). Hence the image $\theta(\partial S)$ is linked (Fig. 6) with the r-axis L , with linking number one. (Adding a point at ∞ to compactify I_1^4 and get the usual linking number of closed curves, either half of the $p = 0$ plane will serve as a spanning surface for L , to prove this technically.) Therefore any spanning surface for $\theta(\partial S)$ will meet L at least once - generically, it will meet L transversely at a finite number of points, odd by link theory , and will not meet 0 . But S is a surface spanning ∂S , and its image in I_1^4 under the same construction is a surface $\theta(S)$ spanning $\theta(\partial S)$. Thus S has at least one point where the expansion of V begins with rx^4 , $r \neq 0$; that is, a cusp point. (Cusp <u>geometry</u> requires that $\theta(S)$ meet L transversely, which generically it does: we are using the transversality principle again, plus the subtle results - notably the preparation theorem - which imply the equivalence of versal unfoldings and hence the possession of particular geometry: see [19] for a long account aimed at the general scientist.)

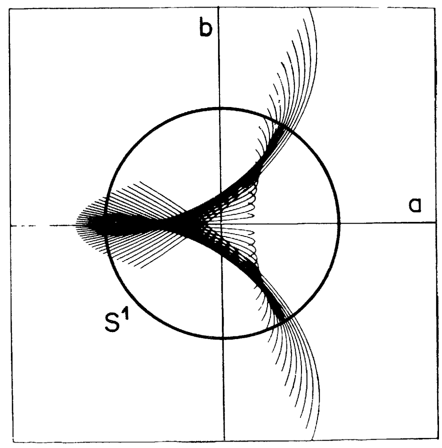
I have described the above argument for its accessibility (the reader unacquainted with linking numbers can probably here intuit the fact and implications of linking) particularly to those acquainted with [18] or [19], in which Thom's theorem is discussed in a closely related way. Furthermore it suggests the promise of a generalised approach of the same type, using deeper relatives of the 'linking number' invariant, as well as allowing one here to 'see' the necessity of a cusp. (But note that the subdivision of I_1^4 into the seven pieces $\{p > 0\}$, $\{p < 0\}$, $\{p = 0 < q\}$, $\{p = 0 > q\}$, $\{p = q = 0 < r\}$, $\{p = q = 0 > r\}$, $\{p = q = r = 0\}$ apparent in Fig. 6 is the beginning of the natural stratification of the space of smooth families of functions, at whose mention we are told [26§13.8.2] that "even the most sophisticated physicist or engineer will prefer to give up". Should it, therefore, have been included in papers addressed to biologists and economists?

The depth of Thom's theorem is significant, not as the misdirection alleged in [26§7], but as more of a barrier to communication than Zeeman in his more optimistic moments is inclined to recognise.)

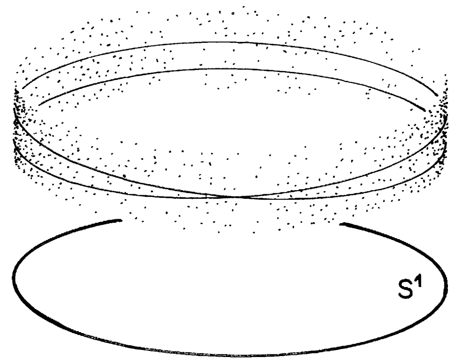
However, Thom's theorem does not permit us here to suppose the state space X one-dimensional. Nor indeed does Zeeman make the false step of assuming we can, which is an invention introduced by Sussmann and Zahler (see above) into their account of [33]. Zeeman does, at the end of [33§8], explain how in the neighbourhood of a cusp we may reduce X effectively to one dimension for many purposes, but only after deducing that the cusp is there. This local result cannot always be globalised. For example, consider the following family of functions on the plane (state variables (x,y), controls (a,b)):

$$V_{ab}(x,y)=x^2+y^2+x^3-3xy^2+3ax+3by \; ,$$

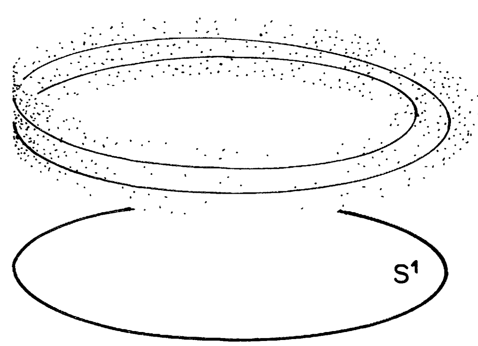
a 2-dimensional family with bifurcation set visible as cusped folds in Fig. 7a. This is stable, generic etc. and obeys the conclusions of Thom's theorem: the only singularities are folds and cusps. Around any point the state space $X = R^2$ can be reduced to



(a) slice of the elliptic umbilic (from [30]).



(b)



(c)

Fig. 7

one dimension plus an 'inessential variable' but this cannot be done globally. Over the unit circle S^1 in the (a,b) plane there is a double loop of equilibria (saddles) which compressed to one state dimension is forced to cross itself* as in Fig. 7b. (If 'states' could be defined in a Möbius strip as in Fig. 7c, it could be one-dimensional over points of S^1 - but such a state 'bundle' cannot be extended over the disc.) In fact no 2-dimensional generic bifurcation geometry with a fold loop that has an odd number of cusps can be globally realised with a one-dimensional state variable. (I do not know of a published general upper bound on the state dimension needed for r controls. For $r = 1$, it is clear that two state variables suffice. For $r = 2$, taking three avoids self-intersection problems, and even two may be enough.)

I have been unable to find an elementary proof, comparable to the one above, for the presence of a cusp in the general case $\dim X = n$: all approaches I have considered have thrown up enough topologists' technicalities to render proofs lengthy and unperspicuous in an account for nonspecialists. At the level of this paper it is more convenient, having established the point at issue under the restriction $\dim X = 1$, to leave the n-dimensional case aside : for a general proof, see [36]. At a more technical level, there is a clear need for development of systematic and general methods for the 'topological boundary value problems' presented by catastrophe theory. Recent work by du Plessis [15] is clearly relevant, though concerned with individual maps with a fixed manifold as domain. Here our concern may be described either as with the singularities of the catastrophe map (the global topology of whose domain, the equilibrium

*) The same dimensional problems force the apparent self-intersections in pictures like those mislabelled in the incompetent Newsweek article (19 Jan. 76) quoted in all of [25, 26, 27], so that such pictures can be misleading if not considered with care.

set, is not known a priori) or with those of the family of maps $V_c : X \to R$ for various controls c .

4. ONLY ONE CUSP?

The boundary conditions clearly do not imply that there cannot be, say, three cusps (though by the linking number argument above for dimX = 1 , like the general proof in [36], parity clearly forbids exactly two). Hence if transversality were only an approximation principle , saying no more than that any 'true' surface can be arbitrarily well fitted by enough cusps, the "ad-hoc hypothesis" [26§7] that there is only one would be as unreasonable as Sussmann and Zahler wish it to appear. However, as a natural 'semi-local' working hypothesis it has, like transversality, solid precedents in successful physics. When Eddington speaks of lines missing "altogether" as in §2 above, he does not mean that they fail by a microunit to coincide. Landau and Lifshitz clearly do not expect to find common zeros of A and B scattered like dust in the (P,T) plane (though no laws are overthrown by occasionally finding two close to each other.) In common scientific experience, most isolated zeros are not merely technically but comfortably isolated: clustering is possible, but gives almost as much of a feeling "now why should that be?" as exact nontransversality, partly because it is so close to it. Most scientists would feel that any one of Figs 27A,B,C of [26], here reproduced as Fig. 8, required some special explanation, despite the way that C is (just) stable and B has only folds. (Contrary to [26§13.4], B is not "nice and 'generic'". Global structural stability is generic by the fullest statements of Thom's theorem if the catastrophe manifold can have, as here, only finitely many layers. The structure of B is perturbed, in a way not removable by equivalence, by an arbitrarily small movement of the vertically stacked folds to separate bifurcation points.)

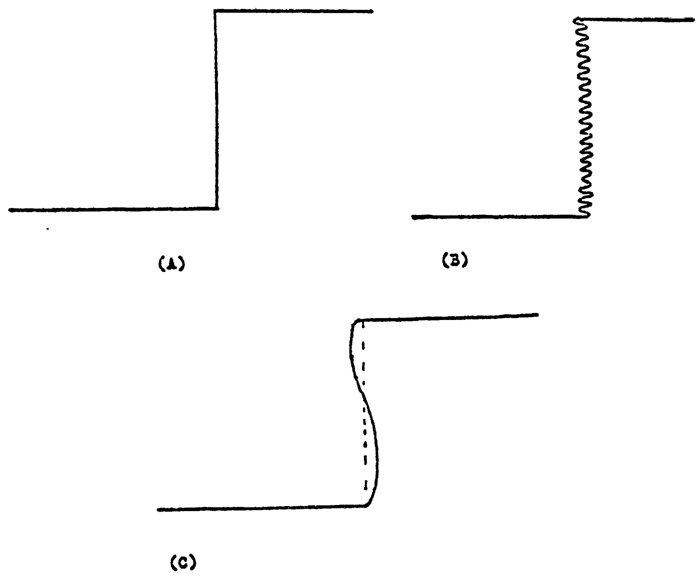


Fig. 8

In close analogy, consider the vibration and electronic spectra of crystals. Physicists initially took it for granted that such spectra had no singularities until these began to appear in the (difficult) exact calculations made for simple models, with important effects on derived thermodynamic quantities like specific heats. Their surprise was dissipated by van Hove [8], who used the Morse inequalities to show that singularities of various types topologically must be present in at least certain specified numbers - they are not accidents of particular crystals or models. This work could give neither the positions of the singularities, nor any upper bound on their number. But a assuming that there are only as many as van Hove's work requires (and that, thanks quite exactly to the transversality principle, they have the nondegenerate types appropriate) has become a standard, and commonly successful initial hypothesis in investigating the physics of a particular crystal. Is then the explicit use of a similar 'minimal singularity' hypothesis unreasonable in biology? Perhaps Zeeman should have listed it separately in [33] from the hypothesis of his potentials being a smooth family, or at least expressed both simultaneously as mathematical aspects of his "Hypothesis II. Continuity", but this is a minor point of style. (Euclid's and Bourbaki's 'Elements', and many journal papers, state all hypotheses very technically at the start, to the despair of readers unfamilar with the kind of statement involved. Hilbert's axiomatisation of Euclidean geometry - the first complete one, after millenia of geometric work - is only comprehensible by working back from the familiar pictures. In teaching, or in communication with nonspecialists, there are advantages in adducing hypotheses as the need for them becomes clear.) Certainly in the large one must expect to often find several cusps - indeed Zeeman's gastrulation model [33] involves multi-cusp structure - but we are here considering the local 'elements' not the global 'compounds' [17] of morphological change, where 'local' means in scientific practice not 'arbitrarily small' but 'at the level of distinguishable phenomena'. If we follow the physicists' lead in expecting cusps usually to be comfortably isolated unless there are definite reasons otherwise, we can use the same ideas in discussing a single free-energy-minimising 'event', whether this is a phase transition or a biological frontier formation.

Some caveats about Zeeman's biological wave model are discussed in [19], and I would certainly oppose his use of 'theorem' as a word applying to matter - Pythagoras's theorem applies to triangles, but only guides us in thinking about rods. This is Thom's point, quoted in §2 above, about mathematics having no right to dictate to reality. (A better case could be made, on precedent, for the word 'law'. For example "Resistance obeys Ohm's Law, except for non-ohmic resistances" is a much more important fact in practice than its tautologous nature might suggest.) The choice of words is definitely unreasonable in Zeeman's "Corollary 1. Initially, when the frontier first forms, it is moving at constant speed." [33 §9] as a way of stating that it has a speed immediately which is neither zero nor infinite. (The latter is not merely a geometric possibility, though nongeneric for cusps; in the 'fast-jump' limit for delay convention it happens stably for the discontinuity of Fig. 9, barred by Zeeman's Hypothesis II, in which asymptotic speed goes - invariantly - as $(t-t_o)^{-\frac{1}{2}}$ after onset at t_o .) A strict adherence to 'perfect' delay makes initially infinite acceleration a consequence of the invariant infinite curvature of the cusp at its tip, so in the idealisation where the frontier has a point origin the velocity is, invariantly, <u>not</u> constant there. Perfect delay is at its most physically dubious there in any case [19], so that this idealisation becomes least useful.

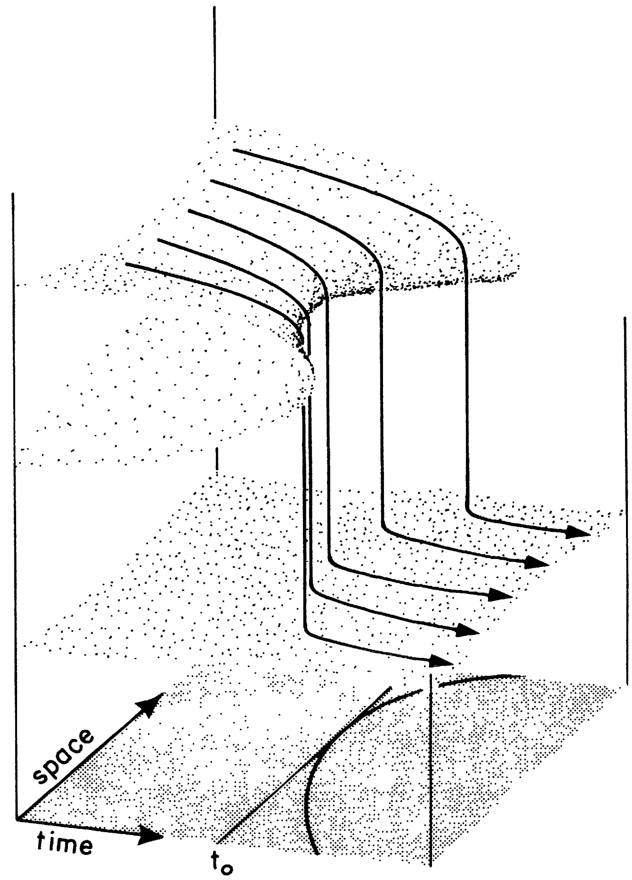


Fig. 9

As such points illustrate, Zeeman's work is no more entirely without flaw than Sussmann and Zahler's is without value. In particular, as regards Zeeman's nerve impulse model [32] some of the discussion in [26] is both excellent and salutary. (Indeed, I presented some of the same strong objections, and some others at the 1975 Mathematical Biology conference at Oberwolfach ; Stewart [22] remarked in the same year on its lack of experimental success. A detailed critique of both the model and its treatment in [26] is in preparation by Woodcock and Stewart. I can find no reference since then to justify calling it "considered by the catastrophists themselves to be especially successful" [26§1] though it has remained important for its ideas and - not least - for the questions clarified by the manner of its failure. There is better evidence that catastrophe theory is a discipline with internal criticism of errors than that "untruth is transformed into truth by repetition" : indeed, the "good example" of the latter given in [26§13.8.4] is an incorrect 1972 statement by Abraham of which, however, no repetitions are cited. If Zeeman were to correct all misstatements about his work published by those not collaborating with him, as footnote 12 of [26] seems to demand , his overload of unproductive writing would include some weeks' labour on [26].) But the light of the nerve impulse model analysis in [26] is somewhat hidden under a bushel of incorrect allegations of gross mathematical error and/or dishonesty in Zeeman's professional field of differential topology, with such terms of scientific debate as "sorcery", "orgy", "hocus-pocus", "sleight-of-hand", "incantations" and "catastrophe theory myth", delivered from the pulpit of the professor of mathematics. To borrow a phrase from [26§7], "most readers will be so intimidated by such words that the mathematical incorrectness will go unnoticed."

5. BIBLIOGRAPHY

[1] R. Abraham & J.W. Robbin, Transversal Mappings and Flows, Benjamin, N.Y. 1967.

[2] V.I. Arnol'd, Wave front evolution and equivariant Morse Lemma, Comm. Pure Appl. Math. 29 (1976) 557-582.

[3] J. Chazarain, Solutions asymptotiques et caustiques, in Rencontre de Cargèse sur les singularités et leurs applications (ed. F. Pham) publ. Math. Dept. Univ. Nice 1975.

[4] R. Courant & H. Robbins, What Is Mathematics?, Oxford University Press 1941.

[5] N.F. Dixon, On the Psychology of Military Incompetence, Johnathan Cape, London 1976.

[6] A.S. Eddington, Gravitation and the Principle of Relativity, Royal Institution Discourse 1918.

[7] P.J. Holmes & D.A. Rand, The bifurcations of Duffing's equation: an application of catastrophe theory, J. Sound Vib. 44 (1976) 237-253.

[8] L. van Hove, The occurrence of singularities in the elastic frequency distribution of a crystal, Phys. Rev. 89 (1953), 1189-1193.

[9] C.A. Isnard & E.C. Zeeman, Some models from catastrophe theory in the social sciences, in Use of Models in the Social Sciences (ed. L. Collins), Tavistock, London 1975.

[10] J.J. Kozak and C.J. Benham, Denaturation; an example of a catastrophe I, Proc. Nat. Acad. Sci. USA 71 (1974), 1977-1981. (The 'review article' [31] was written without awareness of parts II, III in J. Theor. Biol. 63 (1976) 125-149, 66 (1977) 679-693 which had already dealt with those objections in [31] not dependent on misrepresentation.)

[11] H. Kurland & J. Robbin, Infinite codimension and Transversality, in <u>Dynamical Systems</u> - <u>Warwick</u> 1974 (ed. A. Manning), Lecture Notes in Math. 468, Springer, Berlin 1975, 135-150.

[12] L.D. Landau & E.M. Lifshitz, Statistical Physics, 2nd English edition (trans Peierls & Peierls), Pergamon, Oxford 1959.

[13] K. Lorenz, <u>On Aggression</u>, Bantam Books, N.Y. 1967.

[14] R.J. Magnus & T. Poston, On the full unfolding of the von Kármán equation at a double eigenvalue, Math. Report 109, Battelle, Geneva 1977.

[15] A. du Plessis, Maps without certain singularities, to appear in Comm. Math. Helv.

[16] T. Poston, Au courant with differential equations, Manifold 18 (1976) 6-9.

[17] T. Poston, The elements of catastrophe theory <u>or</u> The honing of Occam's razor, to appear in <u>Transformations</u>: <u>Mathematical Approaches to Culture Change</u> (eds K.L. Cooke and A.C. Renfrew), Academic Press, N.Y. 1978.

[18] T. Poston & I.N. Stewart, <u>Taylor Expansions and Catastrophes</u>, Research Notes in Math. 7, Pitman, London & S. Francisco 1976.

[19] T. Poston & I.N. Stewart, <u>Catastrophe Theory and its Applications</u>, Pitman, London & S. Francisco 1978.

[20] M. Spivak, <u>Calculus on Manifolds</u>, Benjamin, N.Y. 1965.

[21] S. Smale, Differentiable dynamical systems, Bull. Amer. Math. Soc. 73 (1967) 747-817.

[22] I.N. Stewart, The Seven Elementary Catastrophes, New Scientist 68 (1975), 447-454.

[23] I.N. Stewart, Catastrophe Theory, Special Supplement to Britannica Book of the Year 1977.

[24] H.J. Sussmann, Some properties of vector fields that are not altered by small perturbations, J. Diff. Eq. 20 (1976) no 2, 292-315.

[25] H.J. Sussmann, Catastrophe theory - a preliminary critical study, PSA 1976 vol. 1 (eds. F. Suppe & P.D. Asquith), Phil. Sci. Assoc., East Lansing, Michigan 1976.

[26] H.J. Sussmann & R.S. Zahler, Catastrophe theory as applied to the social and biological Sciences : a critique, to appear in Synthèse. All quotations are from the widely circulated preprint, Rutgers Univ. Math. Dept., Feb. 1977.

[27] H.J. Sussmann & R.S. Zahler, Catastrophe theory: mathematics misused, The Sciences 17 no 6 (1977) 20-23.

[28] R. Thom, Answer to Christopher Zeeman's reply, in Dynamical Systems - Warwick 1974 (ed. A. Manning), Lecture Notes in Math. 468, Springer, Berlin 1975, 384-389.

[29] R. Thom, Structural stability, catastrophe theory, and applied mathematics, SIAM Review 19, 189-201, 1977.

[30] A.E.R. Woodcock & T. Poston, A Geometrical Study of the Elementary Catastrophes, Lecture Notes in Math. 373, Springer, Berlin 1974.

[31] R.S. Zahler & H.J. Sussmann, Claims and accomplishments of applied catastrophe theory, Nature 269 (1977), 759-763. Letters pointing out many errors, misquotations etc. are in ibid. 270 (1977), 381-384, 658, and a reply by Zahler with some new ones in 271 (1978), 401.

[32] E.C. Zeeman, Differential equations for the heartbeat and nervous impulse, in Towards a Theoretical Biology IV (ed. C.H. Waddington), Edinburgh Univ. Press, 1972, 8-67, Dynamical Systems (ed. M.M. Peixoto), Academic Press, N.Y. 1973, and [35] 81-140.

[33] E.C. Zeeman, Primary and secondary waves in developmental biology, AAAS 1974, Some Mathematical Questions in Biology VIII, Lectures on Mathematics in the Life Sciences 7, Amer. Math. Soc., Providence R.I. 1974, 69-161, and [35] 141-233.

[34] E.C. Zeeman, On the unstable behaviour of stock exchanges, J. Math. Econ. 1 (1974) 39-49, and [35], 361-371.

[35] E.C. Zeeman, Catastrophe Theory: Selected Papers (1972-1977), Addison-Wesley, N.Y. 1977.

[36] E.C. Zeeman, A boundary value problem involving cusps, to appear.

See also I.N. Stewart and T. Poston, The Catastrophe Theory Controversy, to appear in the New Scientist, and the September 1978 special 'Catastrophe Theory' issue of Behavioral Science, where further arguments for catastrophe modelling are variously developed.