COLLOQUE DE CERISY

Logos et Théorie des Catastrophes

A PARTIR DE L'ŒUVRE DE RENÉ THOM

Actes du colloque international de 1982
sous la direction de
JEAN PETITOT

P A T I Ñ O

1988

LE DÉPASSEMENT INTERNE DES PARADIGMES DE LA PHYSIQUE CLASSIQUE

par Michael BERRY

RÉSUMÉ

M. Berry se propose de critiquer certaines composantes de l'image vulgarisée de la physique classique et, ce faisant, de remettre en question certains de ses « dogmes ».

Le premier « dogme » est celui du déterminisme newtonien-laplacien. Il est remis en cause par le développement de la dynamique qualitative depuis Poincaré et, en particulier, par l'étude de systèmes dynamiques qui sont mathématiquement (i.e. idéalement) déterministes tout en étant physiquement (i.e. concrètement) chaotiques et imprédictibles. La possibilité de telles instabilités avait été anticipée par Maxwell, que Berry cite longuement. Il est facile de montrer que le mouvement d'un « gaz » de billes élastiques est réellement imprédictible dans tous les sens pratiques du terme. Les oppositions classiques entre le hasard et la nécessité ou entre l'ordre et le chaos sont donc désormais obsolètes.

Le second « dogme » analysé par Berry est celui de la différentiabilité selon lequel, aux sauts catastrophiques près, les évolutions soumises à des lois exprimées dans des équations doivent être différentiables (« lisses »). Or il existe des structures « fractales » en physique classique. Par exemple celle du mouvement brownien ou celle des fluides au voisinage d'un point critique (dans ce dernier cas une hiérarchie de fluctuations possédant la propriété d'auto-similitude rend impossible la différenciation entre les niveaux respectivement microscopique et macroscopique).

Le troisième « dogme » est enfin celui du réductionnisme. Berry le dénonce en prenant des exemples dans la théorie des caustiques. L'optique géométrique est une approximation asymptotique de l'optique ondulatoire pour des longueurs d'ondes $\lambda \to 0$. Dans un processus de focalisation de la lumière, les caustiques sont — géométriquement — les enveloppes des rayons lumineux. Leurs singularités structurellement stables sont données par la T.C. Mais — ondulatoirement — à chaque singularité est associé un pattern caractéristique d'interférences appelé « catastrophe de diffraction ». Ces catastrophes de diffraction sont optiquement dominantes et, pourtant, elles n'avaient fait jusqu'ici l'objet d'aucune attention particulière. Il est plus satisfaisant de les penser morphologiquement à partir de leurs singularités organisatrices plutôt que comme solutions des équations de Maxwell. C'est en ce sens que l'on peut remettre en cause le réductionnisme.

Il s'agit-là sans doute d'un phénomène général. Le remplacement d'une théorie « profonde » par une théorie « réduite » conceptuellement plus simple semble rencontrer des obstructions quand la théorie réduite possède des *singularités*.

J.P.

BREAKING THE PARADIGMS OF CLASSICAL PHYSICS FROM WITHIN

by Michael BERRY

I. — Introduction

Physics is a successful science in the sense that its models have sufficient conceptual depth to map its subject-matter—the physical universe—with great accuracy, not only on the human scale but on large and small scales far removed from our direct perceptions. This was abundantly clear even in the last century when physics was still in the phase we now call 'classical'. Because of this very success it was natural that sciences which were slower to develop would take methodological inspiration from physics.

My contention is that such imitations may have proceeded from a dangerously restricted notion of what physics is; embodied in what might be called a 'vulgar philosophical model' of that science. I plan to discuss three ingredients of this vulgar philosophical model and show them to be inapplicable even within classical physics (not to mention relativity, quantum mechanics or any other part of what is still called—ridiculously—'modern physics'). At the end I shall describe a suggestion which emerges from the discussion, concerning the place of singularities in physics.

In no way do I wish to be misconstrued as attacking other disciplines—economics, sociology, etc.—about which I know next to nothing. Many practitioners of the so-called soft sciences are well aware of the dangers of naive philosophical modelling. Rather do I wish to liberate classical physics from restricted ideas arising from lack of appreciation of its richness ans subtlety.

II. — The Determinist Fallacy

A. A violent order is disorder; and
B. A great disorder is an order. These
Two things are one[1].

Newtonian mechanics is classical physics par excellence. It gives mathematical rules for calculating the behaviour of particles from a knowledge of their positions and velocities at some initial instant and the forces acting on them. Thus the old idea of causality (« the past determines the future ») was combined with new techniques of mathematical precision into an intoxicating brew giving rise to the opinion that the world is predictable. Laplace even conceived of a superhuman being with knowledge of the instantaneous condition of all the bodies in the universe and able to integrate Newton's equations with lightning facility, who would thereby discover the future (and past) course of events for the whole universe.

There is indeed good evidence to support this opinion, especially in astronomy. Eclipses can be predicted centuries in advance, planets move with high accuracy in their Kepler ellipses, and the orbits of spacecraft can be computed so precisely that they hit the moon within a few kilometres of their target. On Earth, the sinusoidal motion of a particle moving under linear restoring force, becomes, in its practical realisation as a simple pendulum, the very epitome of predictability ("as regular as clockwork").

But some mechanical systems behave quite differently. Just think of a pinball machine, in which a metal sphere bounces elastically amongst an array of cylindrical posts or nails, or of the myriad ceaselessly-colliding molecules in air. These are causal and moreover well described by Newtonian laws. Their motion is however very far from predictable, and better described as 'chaotic' or 'random'.

What is the difference between planets and pinballs? How can very similar equations describe such totally different behaviour? To explain the crucial concept I can do no better than quote from an essay on 'Science and free will'[2] written by Maxwell in 1873, which foreshadowed the revolutionary dynamical studies of Poincaré ten years later and also the whole contemporary explosion of research in 'qualitative dynamics':

> Much light may be thrown on some of these questions by the consideration of stability and instability. When the state of things is such that an infinitely small variation of the present state will alter by only an infinitely small quantity the state at some future time, the condition of the system, whether at rest or in motion, is said to be stable; but when an infinitely small variation in the present state may bring about a finite difference in the state of the system in a finite time, the condition of the system is said to be unstable.
>
> It is manifest that the existence of unstable conditions renders impossible the prediction of future events, if our knowledge of the present state is only approximate, and not accurate.
>
> It has been well pointed out by Professor Balfour Stewart that physical stability is the characteristic of those systems from the contemplation of which determinists draw their arguments...

Maxwell develops the idea that instabilities are related to *singularities.* This is only one of the mechanisms by which according to our present understanding dynamical systems may be unstable, but in the context of this meeting his remarks are well worth repeating:

> In all such cases [of instability] there is one common circumstance,—the system has a quantity of potential energy, which is capable of being transformed into motion, but which cannot begin to be so transformed till the system has reached a certain configuration, to attain which requires an expenditure of work, which in certain cases may be infinitesimally small, and in general bears no definite proportion to the energy

developed in consequence thereof. For example, the rock loosed by frost and balanced on a singular point on the mountain-side, the little spark which kindles the great forest, the little word which sets the world a fighting, the little scruple which prevents a man from doing his will, the little spore which blights all the potatoes, the little gemmule which makes us philosophers or idiots. Every existence above a certain rank has its singular points: the higher the rank, the more of them. At these points, influences whose physical magnitude is for small to be taken account of by a finite being, may produce results of the greatest importance...

If, therefore, those cultivators of physical science from whom the intelligent public deduce their conception of the physicist, and whose style is recognised as marking with a scientific stamp the doctrines they promulgate, are led in pursuit of the arcana of science to the study of the singularities and instabilities, rather than the continuities and stabilities of things, the promotion of natural knowledge may tend to remove that prejudice in favour of determinism which seems to arise from assuming that the physical science of the future is a mere magnified image of that of the past.

The extreme unpredictability of unstable systems is well illustrated by the important example of molecules in air, whose motion is causally governed by Newton's equations describing the deflection during a collision under the action of intermolecular forces. Imagine that all the positions and velocities have been measured at some instant (never mind how!) and consider the subsequent motion of one molecule. The initial condition can never be measured exactly an I suppose the inevitable inaccuracy to be embodied in an angular error δ_0 in the direction of motion.

Collisions will magnify this uncertainty, and it is not hard to show that if the molecules have radius R and mean separation ℓ then each collision will increase δ by a factor ℓ/R, so that after n collisions the angular error is

$$\delta_n = \left(\frac{\ell}{R}\right)^n \delta_0$$

For air, ℓ/R is about 10, and this means that one decimal digit of accuracy in our knowledge of the direction of motion is lost with every collision. If $\delta_0 = 0.00\ldots$ n zeros... 1, then after n collisions δ_n is of order unity so that the molecule's direction is uncertain by about a right angle and this corresponds to a complete loss of predictability.

The dramatic growth in uncertainty arises from the exquisite sensitivity to collisions of convex bodies. But does this example really demolish Laplace's view that *in principle* the motion could be determined? Imagine now that δ_0 were exactly zero, i.e. that we really did know the initial conditions *exactly,* in defiance of imperfections of measuring apparatus, of quantum-mechanical uncertainty, of the impossibility of representing general data with infinite accuracy in a finite space... Surely then we could predict the motion! But now another difficulty arises. We have assumed the system of gas molecules to be *isolated,* so that the only forces are internal. This could be achieved by putting the gas in a box whose walls screen out all external influence. But that is impossible! There is one force which cannot be screened, namely gravity. Gravity is a weak force, so it might be that the predictability of the system is little compromised by the gravitational effect of external matter. But this is not the case. For a dramatic example, consider the weakest disturbence imaginable, namely the gravity of an electron at the observable limit of the universe, acting on our gas molecules, and ask for the number n of collisions it takes for this to make the molecular motion unpredictable in the same sense as before (direction uncertain by a right angle). The answer (easy to calculate) is...n = 50. This number of collisions takes place in 10^{-10} seconds, so that the microscopic motion of a gas is truly unpredictable in every practical sense.

Mathematically, the causal aspect of dynamics lies in the fact that theory gives a *transformation* between conditions at different times, and it is easy to construct abstract transformations, not intended as explicit physical models, to illustrate very clearly the difference between predictable and unpredictable evolution. Let the 'state' be represented by a number x between 0 and 1 and let 'dynamics' be represented by a transformation T between x_n, the state at 'time' n, and x_{n+1}, the state at 'time' x_{n+1}. Thus

$$x_{n+1} = Tx_n.$$

As an example of predictable evolution, consider the squaring transformation

$$Tx = x^2.$$

This rapidly makes x smaller (if $x_0 = \frac{1}{2}$, $x_8 \approx 10^{-77}$), and the system evolves stably with utter predictability, to x = 0, irrespective of the initial state.

As an example of unpredictable evolution, consider the doubling transformation

$$\begin{aligned} Tx &= \text{Fractional part of } 2x \\ &= 2x \text{ if } x < \frac{1}{2} \text{ and } 2x - 1 \text{ if } x \geqslant \frac{1}{2} \end{aligned}$$

The effect of this can be best expressed by writing x_0 in binary notation, i.e.

$$x_0 = 0.a_1a_2a_3...$$

where the a's are zeros and ones. Then T moves the point one place to the right, so that

$$x_n = 0.a_na_{n+1}...$$

Although simply expressed, the evolution of x is highly unstable. To see this, consider two initial conditions x_0 which are closely similar in the sense that their first n binary digits are the same. Their evolutions $x_1...x_n$ will be similar, but beyond x_n their fates are completely unrelated. Another way to regard this transformation is to consider it as a rule for generating *all possible cointossing sequences* a_1 a_2... (0 for heads, 1 for tails), from the successive transformations of a corresponding x_0. Therefore the doubling transformation, a causal rule uniquely associating antecedents with consequents, produces sequences indistinguishable from that exemplar of randomness, the unbiased coin-toss!

What these examples show is that having a causal model of a system, in the form of causal equations precisely governing its evolution, in no way guarantees that the system will be predictable.

The general situation is even more complicated. I have considered the motion of a given type of system as being either predictable or chaotic. But in recent decades it has been realised that complete predictability and complete chaos are in exceptional limiting cases: 'most' dynamical systems partake of both sorts of behaviour, in the sense that some orbits are predictable and some are chaotic, depending on initial conditions. Moreover the different types of orbit are mixed in a most intimately complex way: as the rational numbers are distributed among the reals. An example is the motion of bodies in the solar system. If we ignore the forces between planets and imagine each to be acted on only by the sun's gravity, then the orbits are the completely predictable Kepler ellipses already discussed. But the planets do interact weakly, by mutual gravitation, and when this is taken into account it is found that while most orbits remain stable, there arise narrow zones of instability, such as are manifested by the 'Kirkwook gaps' in the asteroid belt between Mars and Jupiter, which occur near radii where particles would orbit the sun with a period in resonance with that of Jupiter.

This discovery that simple equations can have such marvellously complicated solutions is a good thing for scientists to keep in mind. It helps us to defend ourselves against the accusation that in seeking general laws, expressible as brief scribbles on a piece of paper, we are doing violence to Nature's 'immeasurable' richness.

By way of summary, I end this section by pointing out that the ancient oppositions of chance and necessity, fate and luck, order and chaos, are not really oppositions at all: even the narrow conceptual world of old-fashioned Newtonian mechanics encompasses all varieties of predictability and unpredictability.

III. — The Smoothness Prejudice

On the point of a single hair a whole Buddha land may be seen.

Now we turn to another legacy from Newtonian physics, or rather from the differential calculus invented to enable its laws to be formulated as equations. It is a condition of applicability of calculus that functions (e.g. the position of a body at different times, or the temperature of a fluid at different places) depend *smoothly* on their variables. When applied to graphs, smoothness means that any curve representing the variation of a physical quantity gets more and more like a straight line when looked at on successively finer scales—it gets boring under magnification.

The spectacular success of applications of smooth functions led to the prejudice that smooth evolution (possibly punctuated by catastrophic jumps) is the only kind that can be subject to scientific law. In economics, for example, where trade figures are sampled at discrete times, it is tempting to imagine that underlying these figures are smoothly-varying functions of continuous time, whose laws, in the form of differential equations, it is the job of economic science to determine.

But there are many phenomena which cannot sensibly be modelled by smooth functions because more and more structure is revealed under magnification. For example, on all scales of interest in geography, any natural landscape is not smooth—there are bumps upon bumps upon bumps, as is well illustrated by the impossibility of determining the scale of any unlabelled contour map of an area of terrain. Such phenomena have been systematically studied by Mandelbrot[3], who calls the underlying geometric structures 'Fractals'. He shows that fractals have a dimensionality D that need not be an integer; for example a coastline is a curve so convoluted that it has infinite length (on any reasonable extrapolation from measurements on ever-finer scales), although zero area, and in fact can precisely be described with a D between 1 and 2 (often about 1.2).

My point here is that there are also fractals in the very classical physics from which the smoothness prejudice originated. Consider first that paradigmatic Newtonian system of a ball flying through the air. It moves smoothly in a parabola. Well, almost. In reality it is resisted by the friction of the air and blown off-course by the motion of the air. However these effects are easily taken into account by an extra term in Newton's equation. Is the resulting motion smooth? The first answer is yes, because although turbulent air has a hierarchical (and indeed fractal) velocity structure, this has a smallest scale of a few millimetres, below which the air, and hence the ball it blows about, moves smoothly. The second answer is no, because on much smaller, molecular scales the fluctuations in molecular impacts give the ball's centre of mass an irregular motion, superimposed on the smooth gravity-plus-friction-plus-turbulence, and this can be shown to have a fractal character. But this assumes molecular collisions to be sudden events, giving instantaneous impacts to the ball, and the fact that molecular forces vary smoothly with distance means that even this finest irregular motion when looked at in submicroscopic detail is smooth; therefore the third (and, at least at the classical level, final) answer is yes, the motion is smooth.

Note that the ball's motion does have a regime, between the finest scale and the smallest macroscale, where it is reasonable to describe its motion as a fractal. In this problem, however, the nondifferentiability of the path is not very important. But it is not difficult to construct situations where the fractal regime dominates observation. One such case is *Brownian motion,* where the 'air' is water and the 'ball' a small particle of pollen or dye whose motion is observed under a microscope. The scales observed are large compared with the molecular, but unlike the ball the particle is hardly affected by gravity (because of buoyancy) or turbulence (which is unimportant in a microscope cell). This means that motion, dominated by molecular impacts, is irregular on all scales observed. It follows that the trajectory in spece is a statistically self-similar fractal curve with dimension D = 2, while the graph of any coordinate as a function of time is a fractal curve with D = 1.5.

For the next example, forget the ball and consider only the fluid, in particular the distribution of its molecules. On small scales this distribution is irregular and this shows up as random clustering of the molecules. On large scales the clusters are too small to be seen and fluid appears smooth, with a number density of d molecules per unit volume. To express this mathematically, consider a given molecule and imagine it to lie at the centre of a sphere with radius R. If R is small, the number N of other molecules whose centres lie within the sphere will fluctuate as time proceeds, because of the continual formation and dissolution of clusters involving the central molecule. If R is large, the fact that clusters are usually small (typically a few molecular sizes) means that the fluctuations average out and N lies very close to

$$N \rightarrow \frac{4}{3} \pi R^3 d$$

as $R \rightarrow \infty$. The appearance of the factor R^3 signals the fact that, macroscopically, the fluid is three-dimensional: for a two-dimensional fluid, e.g. a layer adsorbed on a solid surface, the factor would de R^2.

Now, this picture of clearly separated micro- and macroscales, characterised respectively by strong clustering and smooth density, fails if the fluid is prepared in what is called the *critical state,* or *critical point* corresponding to a particular temperature $T = T_c$ (and pressure) characteristic of the substance. At temperatures below critical ($T < T_c$) the fluid forms a liquid or a gas, and undergoes a transition between these phases as (for example) the volume of the container is changed. At temperatures above critical ($T > T_c$) the distinction between liquid and gas disappears and no phase transition can be induced. The critical point is a *singularity of thermodynamics,* in the sense that mathematical quantities representing the properties of the fluid show singular behaviour. For example, the density difference $d_{liquid} - d_{gas}$ vanishes as $T \rightarrow T_c$, and the compressibility of the fluid becomes infinite as $T \rightarrow T_c$.

What is happening on the microscopic scale as this singularity is approached is that the molecular clusters—the transient aggregations—are getting larger and persisting longer. In order to perceive the fluid as smooth, it is necessary to go to larger scales. At the critical point itself, there are *clusters of all sizes,* from individual molecules right up to the size of the container, and the macroscopic state is never achieved. This is a *new state of matter* characterised by a self-similar hierarchy of fluctuations. The number N of molecules within the sphere R surrounding a given one never attains the R^3 distribution but instead behaves like

$$N \rightarrow \text{constant} \times R^D$$

as $R \rightarrow \infty$, where D is a constant slightly less than 2 and independent of the substance. Thus in the critical state matter is a 'dust fractal' with dimension D; it never appears smooth. The development of techniques for calculating D (and other exponents characterising the critical state) on the basis of the (classical) theory of statistical mechanics is a triumph of theoretical physics of the past decade.

A final and dramatic example of the usefulness of non-smooth functions in physics is the universe itself, considered as a fluid whose 'molecules' are galaxies. It is well recognised that galaxies form clusters, clusters form superclusters, etc. There is some dispute about how far this hierarchy continues; but the observed distribution of galaxies in the sky is reproduced with great accuracy by a model of galaxies as a dust fractal with dimension $D \approx 1.3$. The agreement holds not only for the correlations between pairs of galaxies but also for the more delicate correlations between triplets and quartets. Why this value of D occurs, and indeed why the gravitational interaction between galaxies should have led to a fractal distribution, is unknown.

IV. — The Reductionist Dogma

> The dimmed outlines of phenomenal things all merge into one another unless we put on the focussing glass of theory and screw it up sometimes to one pitch of definition and sometimes to another, so as to see down into different depths through the great millstone of the world[4].

It is a common practice in science to explain big things in terms of little things. Bulk phenomena, such as the fluidity of water, the hardness of glass, and the conduction of heat, are explained in terms of the geometric arrangement of, and forces between, atoms. Chemical reactions between molecules are explained in terms of interactions between the electrons in their atoms. (Alchemical) reactions between nuclei are explained in terms of the interactions between their neutrons, protons, mesons, quarks... Molecular biology seeks to explain life's processes and structures in terms of chemical interaction and so, ultimately, in terms of quantum mechanics.

Related to this 'reduction of phenomena' is reduction of the laws governing them; this is 'reduction of theories'. One example is the reduction of thermodynamics, which is the theory of phenomena involving heat, to statistical mechanics which is based on the (classical or quantal) laws governing (unspecified) microscopic constituents. Another is the grand synthesis of the separate theories of electricity, magnetism and optics into the single theory of electromagnetism.

I will not deny the spectacular successes of reductionism, nor that the main motive for scientific activity is to understand phenomena with concepts of ever-wider applicability. But there is growing criticism of reductionism as a dogma, and I want to give an example, directly related to the subject of this meeting, for which the clearest and most immediate understanding is not obtained in terms of the deepest and widest theory. The example is the description of the *focusing of light,* and is concerned with that special part of the reduction of optics to electromagnetism which deals with the connection between waves and rays. (The full reduction of optics includes, in addition, the electromagnetic explanations of polarisation, optical rotation and the optical coefficients of matter).

Light emerges from a small source and is focused by a perfect lens. Two pictures are available to describe this focusing. One can think of rays diverging from the source like the spines of a hedgehog, propagating (according to the laws of geometrical optics) in straight lines until bent (refracted) by the air-glass-air interfaces of the lens so as to converge to the focal point through which all rays pass. Or one can think of waves whose crests form spheres diverging from the source like expanding balloons, which on hitting the lens stimulate wavelets from each point whose subsequent superposition reaches a peak of constructive interference near the focus. Of these two pictures, the deeper is that in terms of waves, which can be obtained more directly (although not completely trivially) from the fundamental equations of electromagnetism. The connection is in terms of the light's wavelength λ: the smaller the wavelength, the more accurate is the description in terms of rays. In other words: ray optics is an asymptotic approximation to wave optics, valid in the limit $\lambda \rightarrow 0$.

So far, all is standard reductionism. But to understand focusing more deeply it must be realised that the perfect lens is a very special system, in that it is exceptional for a family of rays to pass through an isolated focal point. A typical family of rays, after being emitted from a source and reflected or refracted by smoothly-curved surfaces, is focused onto a surface rather than a point. This surface is called the *caustic* of the family; it is the surface touched by all rays in the family, or, alternatively, it is the set of all focal points formed by the intersection of neighbouring rays belonging to the family. Caustics are the singularities of ray theory. On caustics, the light intensity as predicted by geometrical optics, which is proportional to the density of rays, is infinite.

Caustics are bright places in a light field. You can see them on the bottom of a swimming-pool, formed by the focusing of sunlight refracted by the wavy water surface; the bright dancing lines are sections of the caustic surface by the flat pool bottom. You can see them on rainy nights when irregular water droplet 'lenses' on your spectacles focus light from a distant street-lamp onto your retina. And you can see them on rainy days in sunlight refracted and reflected by raindrops: each drop emits light focused on a conical caustic surface, and you see, shining brightly, all the drops on whose cones your eyes lie, which themselves form a cone centred on you and which appears, projected into the sky, as the rainbow's arch.

It is a remarkable fact, quite unsuspected by classical optical scientists, that caustics occurring stably (which persist under small disturbances, unlike the ideal point focus) can be classified. Professor Thom[5] has explained how this classification is in terms of catastrophe theory: catastrophes are loci (in a parameter space) of degenerate critical points of functions, and rays, are, by Fermat's principle, maxima or minima (critical points) of the optical travel time function, whose touching (degeneration) occurs on caustics in space. Therefore the hierarchy of curious universal singular forms generated by catastrophe mathematics is also a hierarchy of stable caustics.

The classification proceeds in terms of codimension, which is the number of dimensions that must be explored in order to find the singularity. The simplest catastrophe, with codimension 1, is the fold, where two rays touch, corresponding to a caustic surface in space or a line in the plane (or the rainbow in two-dimensional angular skyspace). Next, with codimension 2, is the cusp, where three rays touch, corresponding to a line in space (where a fold surface is creased) or a point in plane (where a fold curve reverses direction and crosses its tangent). And so on. In order to find higher catastrophes (including those classified by Arnold whose list continues from codimension 6 where Thom's ends) in the three-dimensional observation space it is necessary to vary additional parameters describing, for example, the form of refracting surfaces. Catastrophe optics is a flourishing subject[6,7] and has helped explain several phenomena in which light is focused 'naturally' (e.g. without lens-makers' artifice). The richest catastrophe so far identified in theory and experiment occurs in the caustic focussed by three superposed water waves. It involves thirteen touching rays and has codimension 11.

The most striking aspect of a caustic is the strong intensification of light caused by the coalescence of rays there. But closer inspection reveals a delicate pattern of brilliant lines and spots arising from the interference of waves corresponding to these rays. The easiest non technological way to see these patterns is in the already-mentioned street-light-through-sprinkled-spectacles-on-a-rainy-night. It is natural to expect each caustic to be decorated by its characteristic interference pattern, and this turns out to be the case. Therefore the classification of caustics, which began at the level of geometrical optics, has become a classification of 'diffraction catastrophes' reaching deep into the wave theory. The individual diffraction catastrophes have intricate structures whose case-by-case exploration is now being carried out both theoretically and experimentally.

An important and fundamental aspect of these patterns is that they are the *most easily observed wave effects* at short wavelengths. Not only is the light most intense (because the patterns decorate caustics) but the scale of the patterns is greatest in the sense that the spacings between neighbouring bright lines and spots is greater than corresponding spacings produced by interference far from caustics. In view of this, the wave theory of light might well have been discovered by people making close observations of caustics, and it is a curious anomaly of history that the subject did not develop in this way.

What all this has to do with reductionism is that during a century's intense study of Maxwell's electromagnetic equations, and one and a half centuries of experience in solving wave equations, nobody even appreciated the existence of the hierarchy of diffraction catastrophes, and nothing was known of their detailed structure except in two special cases. The patterns were only obtained (and, with difficulty, incorporated into the mathematical framework of wave theory) with the aid of concepts involving singularities of the supposedly-superseded (reduced) ray theory.

It is of course a matter of taste which elements of any completed theory one chooses to regard as fundamental, but in my view it is more fruitful to consider diffraction catastrophes as organised by the caustic singularities they decorate than to think of them as solutions of Maxwell's equations. Moreover, caustic catastrophes also play a different role in optics, as 'atoms of form' in a more conventional reduction where they act as morphological elements of large-scale networks of connected caustics from undulating extended wavefronts (as in swimming-pools), whose structure is otherwise inexplicable.

My point is that in order to make sense of this whole range of phenomena involving caustics, it is not appropriate to take the conventional reductionist view which would start with Maxwell's equations and work up. It appears, rather, that the fundamental theoretical objects are the catastrophes. These singularities occupy a sort of mesoscale and organise both the 'macroscopic' caustic networks and the 'microscopic' diffraction catastrophes.

V. — Singularities and Nontrivial Reduction

In the last section I explained that the replacement of ray theory by wave theory, as the wavelength λ becomes small, is not straightforward, and this is connected with singularities (caustics) in the reduced theory (geometrical optics). Now I wish to conjecture that this is a general aspect of the reduction of theories related by the vanishing or divergence of some parameter.

As an example of a trivial reduction, where no singularities are involved, consider the replacement of special relativity by Newtonian mechanics as the speed of light c becomes infinitely great in comparison with the particle speed v. The replacement is a straightforward matter of simple algebra, expressed in terms of kinetic energy for a particle of rest mass m:

$$\text{Kinetic energy} = \underset{\substack{\text{(total} \\ \text{energy)}}}{\frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}}} - \underset{\substack{\text{(rest} \\ \text{energy)}}}{mc^2} = \underset{\substack{\text{(classical} \\ \text{energy)}}}{\frac{1}{2}mv^2} + \underset{\substack{\text{(relativistic} \\ \text{corrections)}}}{\frac{3}{8}\frac{mv^4}{c^2} \cdots\cdots}$$

As the first nontrivial example, consider the replacement of statistical mechanics (a microscopic, atomic, theory) by thermodynamics (a macroscopic, continuum, theory) as the scale of observation gets bigger. We saw in section III that in almost every case the replacement presents no difficulties: for example the granular distribution of atoms can be replaced by a fluid with smooth density. But at the critical point $T = T_c$, which is a singularity of thermodynamics, no scale of observation is large enough to reveal the fluid as a continuum, and indeed the critical state of matter is characterised by fractal structure.

In the wave-ray example, where the limit is $\lambda \rightarrow 0$, the point is that although almost everywhere the replacement consists simply in writing the wave as a sum (superposition) of terms corresponding to the geometrical rays, this procedure fails on the caustics, where the replacement, in terms of diffraction catastrophes, is geometrically and analytically more complicated (and interesting). I remark that the wave $\rightarrow$ ray replacement as $\lambda \rightarrow 0$ is mathematically identical to the replacement quantum mechanics $\rightarrow$ classical mechanics as Planck's constant vanishes, and this observation has led to a catastrophe-theoretical rejuvenation of 'semi-classical mechanics' closely similar to that in short-wave optics.

It seems, therefore, that the replacement of a 'deep' theory by a conceptually-simpler 'reduced' theory is likely to be obstructed whenever the reduced theory has singularities. At such places the reduction is nontrivial and associated with interesting geometric structures.

To finish, it is appropriate to point out that in recent years it is Professor Thom above all who has emphasised both the need to study singularities and the dangers of dogmatic reductionism.

NOTES

1. From 'Connoisseur of Chaos' by Wallace Stevens, in 'The collected poems of Wallace Stevens' (Borzoi Books, A.A. Knopf, New York, 1955) p. 215.
2. In 'The Life of James Clerk Maxwell' by Lewis Campbell and William Garnett (Macmillan, London, 1882) p. 434.
3. 'The Fractal Geometry of Nature' by B.B. Mandelbrot (Freeman, San Francisco, 1982).
4. Ref. 2, p. 237.
5. 'Structural Stability and Morphogenesis' by R. Thom (Benjamin, New York, 1975).
6. 'Catastrophe Theory and its Applications' by T. Poston and I. Stewart (Pitman, London, 1978).
7. 'Catastrophe Optics: Morphologies of Caustics and their Diffraction Patterns', by M.V. Berry and C. Upstill in Progress in Optics vol. XVIII 1980, p. 257.