Logic, Methodology and Philosophy of Science IX
D. Prawitz, B. Skyrms and D. Westerståhl (Editors)

ASYMPTOTICS, SINGULARITIES AND THE REDUCTION OF THEORIES

MICHAEL BERRY

H H Wills Physics Laboratory, Tyndall Avenue, Bristol BS8 1TL, UK

1. Introduction

In science we strive to integrate our experiences, observations, and experiments into a single explanatory framework - 'a theory of everything'. Of course this goal has not been achieved, and probably never will be. What we have instead are the partial descriptions provided by biology, chemistry, physics, etc., and, within these, the various subfields such as fluid mechanics and quantum mechanics. The different areas of study do not fit tidily together. Particular difficulties arise when a more general description is supposed to encompass an older, less general, one, usually by providing a microscopic explanation of its principles. It is hoped that a less general theory can thus be 'reduced' to a more general one. But this comfortable picture is often spoilt by certain classes of higher-level, or 'emergent', phenomena which are well described by the older theory but obstinately refuse to emerge from the supposedly encompassing one.

To illustrate the point with a familiar example, consider life. Is it contained in, or implied by, Schrödinger's equation for the 10^{23} electrons and nuclei in an organism, plus rules for incorporating the environment? I suspect that most scientists, especially physicists, would, if pressed, answer yes, but be uncomfortable. The discomfort stems from a dilemma. We know that writing down the Schrödinger equation and gazing at it is not a promising strategy for finding a cure for AIDS, or learning why we do not live for ever. But we feel that invoking something else, outside physics, at a fundamental level, is mysticism. Somehow, life might emerge from physics in some limit (possibly involving increasing complexity), but we have no clear idea how to convert this dream into science.

Of course this problem of reduction has been studied a great deal by philosophers. Sometimes the discussion centres on the conflict between the two views summed up by the terms 'correspondence' and 'incommensurability': in brief, two theories correspond if one can be deduced as a special case of the other, and are incommensurate if their foundations are logically incompatible. My intention here is to present an idea which seems to capture an essential aspect of the problem of reduction of emergent phenomena and which goes some way towards dissolving the antinomy between incommensurability and correspondence, but which has not to my knowledge been considered by philosophers. I will confine myself to reductions of theories within physics, but of course hope that the idea could eventually prove useful in grander contexts such as the reduction of biology to physics (or chemistry).

To begin, realise that theories in physics are mathematical; they are formal systems, embodied in equations. Therefore we can expect questions of reduction to be questions of mathematics: how are the equations, or solutions of equations, of one theory, related to those of another? The less general theory must appear as a particular case of the encompassing one, as some dimensionless parameter - call it δ - takes a particular limiting value. A general way of writing this scheme is

$$\text{encompassing theory} \rightarrow \text{less general theory} \quad \text{as } \delta \rightarrow 0 \qquad (1)$$

Thus reduction must involve the study of limits, that is asymptotics. The crucial question will be: what is the nature of the limit $\delta \rightarrow 0$? We shall see that very often reduction is obstructed by the fact that the limit is *highly singular.* Moreover, the type of singularity is important, and the singularities are not only directly connected to the existence of emergent phenomena but underlie some of the most difficult and intensively-studied problems in physics today.

There is one aspect of the study of limits in physics which has attracted the attention of philosophers, beginning with Berkeley, that I will not be considering here, even though there are interesting and subtle points still to be brought out. This centres on the fact that the limit $\delta = 0$ is always an idealization; in any actual situation, δ is always finite. Instead of discussing this important matter, which involves the relation between the world and our models of it, I shall remain firmly in the realm of theory.

Before proceeding to examples, I must disambiguate an irritating terminological orthogonality. Philosophers consider the less general theory as being 'reduced by' the encompassing theory, because the latter employs principles that are more elementary to explain more phenomena [1]. Physicists, however, find it more natural to think of the reduction as occurring the other way, that is by the more general theory 'reducing to' the less general one as $\delta \rightarrow 0$ because the less general one is a special case (thus the function $\cos\theta$ 'reduces to' 1 as $\theta \rightarrow 0$).

2. Singular limits and emergent phenomena

Here are six examples of the scheme (1) in physics, together with the meaning of the dimensionless parameter δ.

- special relativity → Newtonian mechanics, $\delta = \nu/c$.
- general relativity → special relativity, $\delta = Gm/c^2a$.
- statistical mechanics → thermodynamics, $\delta = 1/N$.
- viscous (Navier-Stokes) flow → inviscid (Euler) flow, $\delta = 1/Re = \eta/\rho a \nu$.
- wave optics → ray optics, $\delta = \lambda/a$.
- quantum mechanics → classical mechanics, $\delta = \hbar/S$.

Here the meaning of the symbols is as follows. ν : speed of body; c: light speed; G: Newton's gravitational constant; m: mass of body; a: typical linear dimension of body; N: number of particles; Re: Reynolds' number; η : viscosity; ρ: density; λ :wavelength; $\hbar$: Planck's constant; S: typical classical action.

Reduction in its simplest form is well illustrated by the first example. Every physics student learns that one form of the connection between the encompassing theory of special relativity and the less general theory of Newtonian mechanics is contained in the 'low speed' series expansion

$$\sqrt{1-\delta^2} = 1 - \frac{1}{2}\delta^2 - \frac{1}{8}\delta^2 + \cdots \tag{2}$$

The left side represents special relativity, and the right side is a convergent Taylor series whose first term represents Newtonian mechanics. Mathematically, special relativity is analytic in δ at $\delta = 0$, so that the limit is unproblematic (the hyper-relativistic limit $\delta = 1$ is singular, but that is a different matter).

My main point will be that this simple state of affairs is an exceptional situation. Usually, limits of physical theories are not analytic: they are singular, and the emergent phenomena associated with reduction are contained in the singularity. Often, these emergent phenomena inhabit the borderland between theories.

To begin, consider the third example, namely the reduction of thermodynamics by statistical mechanics as the number of particles ($N = 1/\delta$) increases to infinity (the 'thermodynamic limit'). Standard arguments [2] involving large-N asymptotics show that for a fluid the thermodynamic equation of state, e.g. the pressure $P(V,T)$ as a function of volume and temperature, can (in principle and to a large extent in practice) be derived from the principles of statistical mechanics and a knowledge of the forces between the atoms. But the reduction runs into difficulty near the *critical point* P_c, V_c, T_c , where the compressibility $\kappa \equiv [-V(\partial P/\partial V)_T]^{-1}$ is infinite. The problem is to find the form of the divergence of κ as $T \to T_c$. This is a power-law, whose exponent is wrongly given by otherwise useful models such as the Van der Waals theory.

The reason for the difficulty is fundamental, and only after a decade of concentrated effort was it clarified, and techniques developed for the correct calculation of 'critical exponents'. Thermodynamics is a continuum theory, so reduction has to show that density fluctuations arising from interatomic forces have a finite (and microscopic) range. This is true everywhere except at the critical point, where there are fluctuations on all scales up to the sample size. Thus at criticality the continuum limit does not exist, corresponding to a new state of matter [3]. In terms of our general picture, the critical state is a singularity of thermodynamics, at which its smooth reduction to statistical mechanics breaks down; nevertheless, out of this singularity emerges a large class of new 'critical phenomena', which can be understood by careful study of the large-N asymptotics.

A particularly vicious example, at the cutting edge of applied mathematics nowadays, is the fourth on the above list, namely the mechanics of a fluid as its viscosity is decreased or its speed is increased (so that δ gets smaller). Exact solution of the Navier-Stokes equation for smooth flow down a pipe, driven by a pressure difference ΔP , predicts that the mass flow rate is proportional to ΔP . For small δ, however, experiment shows a rate close to $\sqrt{\Delta P}$. The reason is that the predicted flow is unstable, and the true flow is not smooth but disorderly, that is, *turbulent*. In turbulence [4-6], instead of viscous dissipation vanishing smoothly as $\delta \to 0$, the dissipation concentrates onto a set of zero measure which is fractal in form. Again the limit $\delta \to 0$ is singular, and out of the singularity emerges an important phenomenon, namely turbulence, whose mathematical nature is still far from understood.

3. Quantum and classical mechanics

Now we come to the examples I shall discuss in most detail - not because they are more fundamental than the others but because they lie closest to my own research interests [7] - namely the reduction of ray theory (e.g. geometrical optics) to wave theory, and (closely related) of classical to quantum mechanics. Here, singular limits abound, even in the simplest problems, as the following example shows.

A wave (of light, sound or water, for example) travelling along the x-axis with speed ν can be represented by

$$\psi = \cos\left\{\frac{2\pi}{\lambda}(x - \nu t)\right\} \tag{3}$$

In the ray limit (where for example geometrical optics provides a consistent and serviceable description of, for example the operation of telescopes and cameras), we have $\lambda \to 0$. But this limit is singular! ψ is non-analytic at $\lambda = 0$, so that it cannot be expanded in powers of λ ; instead, this wavefunction oscillates infinitely fast and takes all values between -1 and $+1$ infinitely often in any finite range of x or t. Only if we consider the wave *intensity*, corresponding to ψ^2, and average over a small interval corresponding to the finite resolution of a detector, do we get the finite and smooth result corresponding to the intensity of the system of parallel rays corresponding to (3); often it is convenient to average over time (reflecting the fact that for light or sound the wave frequency is too high to measure directly):

$$\left\langle\psi^2\right\rangle_t = \left\langle\cos^2\left\{\frac{2\pi}{\lambda}(x - \nu t)\right\}\right\rangle_t = \frac{1}{2} \tag{4}$$

Now consider the superposition of two such waves, with speeds ν and $-\nu$, giving

$$\psi = \cos\left\{\frac{2\pi}{\lambda}(x - \nu t)\right\} + \cos\left\{\frac{2\pi}{\lambda}(x + \nu t)\right\} = 2\cos\left\{\frac{2\pi x}{\lambda}\right\}\cos\left\{\frac{2\pi\nu}{\lambda}\right\} \tag{5}$$

and the time average

$$\left\langle\psi^2\right\rangle_t = 2\cos^2\left\{\frac{2\pi x}{\lambda}\right\} \tag{6}$$

This describes a spatially fixed interference pattern such as that produced by a double slit. Again there is a powerful singularity at $\lambda = 0$. To eliminate it requires an extra average, this time spatial, and then we obtain

$$\left\langle\psi^2\right\rangle_{t,x} = 1 \tag{7}$$

Thus to obtain from wave theory the simple fact that in ray theory two beams of intensity 1/2 add to give intensity 1, with no interference, requires a double average over a mathematically pathological function.

Having seen that interference is associated with a $\cos^2(1/\lambda)$ singularity in the ray limit, we now examine the anatomy of other sorts of wave singularity. An interesting case occurs when waves reach places that rays do not.

Examples are the outside of a glass-air interface within which total internal reflection occurs, the dark side of a rainbow, and the thin layer of air near a hot road in which mirage reflections are seen. In the ray limit, the wave and its intensity are zero, but there are nevertheless waves present, whose amplitude is typically

$$\psi \propto \exp\left\{\frac{-\text{function of } x}{\lambda}\right\} \tag{8}$$

Again this is singular, and cannot be expanded in a power series in λ (all terms are zero).

An important role is played by the *transition* between these two sorts of singularity ($\cos^2\{1/\lambda\}$ and $\exp\{-1/\lambda\}$). (This is somewhat analogous to the transition T through T_c in thermodynamics, for large N.) The transition happens across a *caustic*, which is an envelope of a family of rays (a generalized focal surface in space, or line in the plane), marking the boundary between regions with different numbers of rays. In the simplest case, the regions have two rays and no rays, corresponding to the 'interference' and 'penetration' regimes represented by (6) and (8). A caustic is a collective phenomenon, a property of a family of rays that is not present in any individual ray. Probably the most familiar example is the rainbow. The singularity across a caustic must interpolate between (6) and (8). How this happens was first elucidated by Airy in 1838 as part of an attempt to understand supernumerary rainbows, that is oscillations on the lit side of the bow, in the intensity of light of a given colour. It was necessary for him to invent a new function $Ai(z)$, oscillatory for $z < 0$ and decaying for $z > 0$. In terms of $Ai(z)$, the wave across a caustic has the form

$$\psi = \frac{1}{\lambda^{1/6}} Ai\left\{\frac{Kx}{\lambda^{2/3}}\right\} \tag{9}$$

In this transition, the emergent phenomenon is the fringe pattern associated with a caustic: in the ray limit $\lambda \to 0$, its intensity grows as $\lambda^{-1/3}$, and the spacing of the fringes shrinks as $\lambda^{2/3}$.

Caustics can themselves have singularities, whose classification is the province of catastrophe theory [8]. At such places, the envelope of rays is itself singular. These singular envelopes are decorated with wave patterns ψ whose $\lambda \to 0$ singularities (shrinking fringe spacings and diverging intensities) depend on the geometry of the catastrophe. Such 'diffraction catastrophes' have intricate and beautiful structures [9,10], and constitute a hierarchy of nonanalyticities, of emergent phenomena par excellence. The patterns inhabit the borderland between the wave and ray theories, because when λ is zero the fringes are too small to see, whereas when λ is too large the overall structure of the pattern cannot be discerned: they are *wave* fringes decorating *ray* singularities.

Quantum mechanics is a particular wave theory, whose corresponding ray theory is classical mechanics, and where Planck's constant $\hbar$ plays the role of wavelength λ (through De Broglie's relation $\lambda = 2\pi\hbar/p$ where p is momentum). Its relation to classical mechanics should be through the *semiclassical limit* $\hbar \to 0$. When the limit is not singular, we have the correspondence principle: quantum observables tend to their classical counterparts as $\hbar \to 0$. Usually, though, the limit is singular, and then the correspondence principle, while often a useful guide [7], is too crude to be a substitute for mathematical asymptotics. From the analogy with other sorts of waves we expect that the nonanalyticities and emergent caustic phenomena described above will occur in quantum mechanics, and these have indeed been seen in the scattering of electrons, nuclei and atoms. In addition, the $\hbar \to 0$ limit is enriched by another limit, which is fundamental, namely the *long-time limit* $t \to \infty$.

There are several reasons to study the long-time limit in conjunction with the semiclassical limit:

■ Spectra of atoms and molecules involve the quantized energies of these systems when in stationary states. These are states that persist over infinite time, so their semiclassical study – spectra near the classical limit – inescapably involves $t \to \infty$ too.

■ Experiments on atoms traversing strong oscillating fields begin to probe the combined $\hbar \to 0, t \to \infty$ limit.

■ It is only after infinite time that chaos may occur in the classical orbits. Chaos [11, 12] is unpredictability arising from exponential sensitivity to initial conditions in a bounded region. Therefore any attempt to study how classical chaos is reflected in the semiclassical limit of quantum mechanics ('quantum chaology' [13, 14]) must evidently involve $t \to \infty$ as well.

The essential point is that *the two limits do not commute*: taking the classical limit first, and the long-time limit second, leads to a different result from taking the limits in the reverse order. Such a clash of limits implies a singularity at the origin of the plane with coordinates $\hbar, 1/t$. One way to try to resolve the clash is to take both limits at once, in a controlled way, i.e.

$$\hbar \to 0,\ t \to \infty,\ \hbar t \equiv \tau = \text{constant} \tag{10}$$

In the one case where it has been possible to take the combined limit explicitly [15], for a system whose classical dynamics is trivial, analysis shows that the point $\hbar = 1/t = 0$ is truly a 'dragon's lair' , so singular that the behaviour exhibits a fantastic complexity which depends on the *arithmetic nature* of τ.

When the classical orbits are chaotic, the clash of limits generates some remarkable emergent phenomena. I will briefly describe just one: *the statistics of spectral fluctuations*. Consider a bound quantum system, that is one with a discrete spectrum of energy levels, and ask about the distribution of these levels in the semiclassical limit. The simplest fact about the levels is that as $\hbar \to 0$ they get closer together – their mean spacing is proportional to $\hbar^N$, where N is the number of freedoms. This must happen, because in the classical limit the levels form a continuum. (It is worth pausing to remark that this particular passage to the limit provides a nice illustration of the 'incommensurability' and 'correspondence' approaches to reduction. In the first, it is emphasized that for any finite $\hbar$, however small, the spectrum is always discrete: the classical continuum is never reached, and so cannot be said to be logically contained in the semiclassical limit. On the other hand, when $\hbar$ is sufficiently small the inevitably finite resolution of any spectroscopic measuring device means that the results of all observations will be the same as if the spectrum were continuous, and the correspondence principle can be said to apply.)

Now imagine looking at the set of levels with a microscope [14] whose power is proportional to the mean level density, thus generating a rescaled spectrum consisting of a set of numbers whose mean density remains constant as $\hbar \to 0$. What is the statistical nature of the fluctuations of this set of numbers about its (unit) mean density? The answer is remarkable: apart from trivial exceptions, the fluctuations are *universal* [14, 16], that is, independent of the details of the system and dependent only on whether the orbits of its classical counterpart are regular or chaotic. Paradoxically, the spectral fluctuations are those of a sequence of random numbers (Poisson distribution) when the classical motion is regular, and are more regularly distributed (exhibiting the level repulsion characteristic of the eigenvalues of random matrices) when the classical motion is chaotic. We are beginning to understand this quantum universality [7] in terms of semiclassical asymptotics: it arises from a similar universality in the distribution of long-period classical orbits.

Universality of the spectral fluctuations is a novel qualitative phenomenon emerging from quantum mechanics in the combined semiclassical long-time limit. It was not predicted by analysis of the Schrödinger equation, but was discovered in numerical experiments (and later seen in real experiments) motivated by some physical arguments. Nevertheless Schrödinger's equation does contain it, albeit well concealed behind some very tricky (and incompletely explored) asymptotics.

4. Divergent series

So far, we have considered only the leading-order behaviour in the parameter δ whose vanishing describes how the encompassing theory reduces to the less general one. In those cases where any sort of mathematical treatment was possible, the leading-order behaviour was quite complicated (cf. equation (9)), and this of course reflects the singular nature of the limit. But determination of the leading order is only the first step: a complete treatment requires understanding the series consisting of all the correction terms – usually involving powers of δ. The determination of such series is still in its infancy, but it has been carried out for certain of the simpler problems of wave physics described in §3.

The most important characteristic of such series, and one which almost certainly extends to all series associated with singular reductions, is that they *diverge*. This was one of the factors prompting a re-examination [17] of the mathematics and physics of divergent series. The main results reinforce earlier indications [18] that the divergent tail, conventionally discarded as mathematically meaningless, contains important information in coded form. When decoded, these tails not only enable the function being expanded to be approximated to previously unequalled levels of accuracy but also describe physical effects, associated with the reduction that the asymptotics is attempting to describe, which are qualitatively different from those contained in the leading terms. Examples are the exponentially weak births of rays beyond caustics [19] and the generation of transitions between quantum states [20] in the adiabatic limit of slow driving.

It seems clear that these ideas, and further developments of them, must be involved in any complete description of how the less general theory is embedded in the structure of the encompassing theory.

5. Concluding remarks

Even in what philosophers might regard as the simplest reductions, between different areas within physics, the detailed working-out of how one theory can contain another has been achieved in only a few cases and involves sophisticated ideas on the forefront of physics and mathematics today. This is because in all nontrivial reductions the encompassing theory is a singular perturbation (parameterised by δ) of the less general one. The singularities are reflected in the quantities of the encompassing theory being nonanalytic at $\delta = 0$, and the nonanalyticities describe emergent phenomena in the borderland between the theories. As examples of these phenomena I described thermodynamic critical behaviour in fluids, fluid turbulence, interference patterns decorating optical caustics, and the chaology-dependent statistics of energy-level fluctuations in quantum mechanics.

It should be clear from the foregoing that a subtle and sophisticated understanding of the relation between theories within physics requires real mathematics, and not only verbal, conceptual and logical analysis as currently employed by philosophers. One can hope that these ideas generalize beyond physics (for example to the reduction of biology or chemistry). This would mean that the problem of theory reduction would itself have been 'reduced', to the mathematical asymptotics of singularities. From the evidence so far, the task will be far from easy, and will require the development of new physical ideas and new mathematical concepts and techniques.

Finally, I would be the first to admit that the ideas explored here lack precision in several respects, and have not been presented in their final form. I hope they will benefit from the attention of philosophers.

References

[1] BULLOCK, A. and STALLYBRASS, O. 1977, *The Fontana Dictionary of Modern Thought.* (Collins, London). "Reduction: ... In philosophy ... the process whereby concepts ... that apply to one type of entity are redefined in terms of concepts ... of another kind, normally one regarded as more elementary ...".

[2] TOLMAN, R.C. 1938, *The Principles of Statistical Mechanics* (Oxford, University Press).

[3] WILSON, K.G. 1975, Rev. Mod. Phys. 47, 773-840.

[4] MANDELBROT, B.B. 1982, *The Fractal Geometry of Nature* (San Francisco: Freeman).

[5] FRISCH, U, 1983, in *Chaotic Behaviour in Deterministic Systems*, eds G. Iooss, R.H.G. Helleman and R. Stora, Les Houches Lecture Series XXXVI (Amsterdam: North-Holland) pp665-704.

[6] SREENIVASAN, K.R. and PRASAD, R. 1989, Physica D38, 332-339.

[7] BERRY, M.V. 1991, in *Chaos and Quantum Physics*, eds M-J Giannoni, A. Voros and J. Zinn-Justin, Les Houches Lecture Series LII (Amsterdam: North-Holland).

[8] POSTON, T. and STEWART, I.N. 1978, *Catastrophe Theory and its Applications* (London: Pitman).

[9] BERRY, M.V. and UPSTILL,C.1980, *Progress in Optics* 18, 257-346.

[10] BERRY, M.V. 1990, *Current Science* (India), 59, 1175-1191.

[11] STEWART, I.N. 1989 *Does God Play Dice? The Mathematics of Chaos* (Oxford: Blackwell).

[12] BERRY, M.V. 1990, Proc.Roy. Institution of Gt. Britain, 61, 189-204.

[13] BERRY,M.V., 1989, Physica Scripta 40 335-336.

[14] BERRY, M.V., 1987, Proc.Roy.Soc. A 413, 183-198.

[15] BERRY, M.V., 1988, Physica D, 33, 26-33.

[16] BOHIGAS, O. and GIANNONI, M-J.1984 in *Mathematical and Computational Methods in Nuclear Physics,* eds., J.S.Dehesa, J.M.G.Gomez and A.Polls, Lecture Notes in Physics 209 (N.Y.:Springer-Verlag) pp1-99.

[17] BERRY, M.V. 1991 *Asymptotics, superasymptotics, hyperasymptotics...* in Asymptotics beyond all orders, ed. S. Tanveer (New York: Plenum) in press.

[18] DINGLE, R.B. 1973, *Asymptotic Expansions: their Derivation and Interpretation* (London: Academic Press).
[19] BERRY, M.V., 1989, Publ. Math.of the Institut des Hautes Études scientifique, 68 211-221.
[20] BERRY, M.V. 1990 Proc.Roy.Soc.Lond, A429, 61-72.