ROBERT W. BATTERMAN

THEORIES BETWEEN THEORIES: ASYMPTOTIC LIMITING INTERTHEORETIC RELATIONS*

ABSTRACT. This paper addresses a relatively common "scientific" (as opposed to philosophical) conception of intertheoretic reduction between physical theories. This is the sense of reduction in which one (typically newer and more refined) theory is said to reduce to another (typically older and "coarser") theory in the limit as some small parameter tends to zero. Three examples of such reductions are discussed: First, the reduction of Special Relativity (SR) to Newtonian Mechanics (NM) as $(v/c)^2 \to 0$; second, the reduction of wave optics to geometrical optics as $\lambda \to 0$; and third, the reduction of Quantum Mechanics (QM) to Classical Mechanics (CM) as $\hbar \to 0$. I argue for the following two claims. First, the case of SR reducing to NM is an instance of a genuine reductive relationship while the latter two cases are not. The reason for this concerns the nature of the limiting relationships between the theory pairs. In the SR/NM case, it is possible to consider SR as a regular perturbation of NM; whereas in the cases of wave and geometrical optics and QM/CM, the perturbation problem is singular. The second claim I wish to support is that as a result of the singular nature of the limits between these theory pairs, it is reasonable to maintain that third theories exist describing the asymptotic limiting domains. In the optics case, such a theory has been called "catastrophe optics". In the QM/CM case, it is semiclassical mechanics. Aspects of both theories are discussed in some detail.

0.

In an important paper (Nickles 1973) on reduction in physical theories Thomas Nickles argues that there are two distinct concepts of reduction appearing in the literature. The first is the usual "philosophical" sense which crudely speaking, depends on the deducibility of the reduced theory from the more fundamental reducing theory. It is often suggested that this sort of reduction leads to the explanation of the reduced by the reducing theory. But Nickles notes a second use of "reduction" – more often found in the writings of physicists than philosophers. In this sense the more fundamental theory is said to reduce to the less fundamental (typically superseded) theory in a limiting domain.

This paper is in part concerned with this second, physicists', sense of theory reduction; what Nickles called "reduction$_2$". I agree that this is a legitimate use of the term "reduction", but I claim that its form and plausibility depends crucially on the nature of the relationship between

Synthese **103:** 171–201, 1995.

the two theories in the limiting domain. It has been said, for example, that special relativity (SR) reduces_2 to Newtonian mechanics (NM) in the limiting domain where velocities are small compared with the speed of light. Similarly, it has been said that quantum mechanics (QM) reduces_2 to classical mechanics (CM) in the limit as Planck's constant h (or $\hbar = h/2\pi$) approaches zero. A third case is the limiting relationship between wave optics and geometrical optics. My main motivation for studying intertheoretic reduction is not so much to try to make general claims about the nature of reduction as it is in understanding the particular and peculiar connections and correspondences between certain pairs of theories; particularly, between classical and quantum mechanics. It seems to me that contrary to popular belief, and in stark contrast to the case of SR/NM, there is a very important sense in which no reductive relationship obtains between members of the second two pairs of theories. The reason for the failure is interesting and leads to new physics and interesting philosophy which gets obscured if one accepts a claim of reduction at face value.

The paper begins with a discussion of some philosophical literature on intertheoretic reduction. I focus particularly on a discussion by Fritz Rohrlich concerning a form of reduction involving limiting relations between theories. A strong case can be made for the claim that NM reduces to SR along the lines of Rohrlich's proposal. Following this, I consider the case of wave and geometrical optics and then turn to the quantum/classical case. The same conclusion is not forthcoming concerning these latter two cases.

1.

A paradigm case where a limiting reduction rather straightforwardly does obtain is that of classical Newtonian particle mechanics and the special theory of relativity.[1] In the limit where $(v/c)^2 \to 0$ SR reduces to NM. (I am going to drop the subscript on "reduction" as the kind of reduction discussed will be evident from the context.) We will consider this case first.

As Nickles says, "epitomizing [the intertheoretic reduction of SR to NM] is the reduction of the Einsteinian formula for momentum,

$$p = \frac{m_0 v}{\sqrt{1 - \frac{v^2}{c^2}}}$$

(where m_0 is rest mass), to the classical formula $p = m_0 v$, in the limit as $v \to 0$" (Nickles 1973, p. 182).[2] There is no claim, of course, that every formula of SR will reduce in the limit $(v/c)^2 \to 0$ to a formula of NM. For example, Einstein's famous equation $E = mc^2$ cannot be so reduced since $E = m_0 c^2$ is not a formula of NM. But this merely expresses the fact that SR actually goes beyond NM.

Fritz Rohrlich argues that NM "reduces" to SR (in an attenuated philosophers' sense)[3] because the *mathematical framework* of SR reduces to the *mathematical framework* of NM. Thus, he in effect is claiming that one can demonstrate a "philosophers'" reduction *because* a "physicists'" reduction obtains. Here the idea is that the mathematical framework of NM is "rigorously derived" from that of SR in a "derivation which involves limiting procedures" (Rohrlich (1988), p. 303). The philosophical reduction is attenuated in the sense that he holds that only the mathematical structures of the two theories can be related by this limiting derivational procedure; whereas, "the interpretations and the ensuing ontologies [of the two theories] are in general not so related" (Rohrlich (1988), p. 303). Roughly speaking, for Rohrlich a "coarser" theory is reducible to a "finer" theory in the philosophers' sense of being rigorously deduced from the latter just in case the mathematical framework of the finer theory reduces in the physicists' sense to the mathematical framework of the coarser theory. In the case of NM (the coarser theory) and SR (the finer theory), he expressed this as follows:

(i) $((v/c)^2 \to 0) \lim \mathrm{M(SR)} = \mathrm{M(NM)}$

Rohrlich also speaks of "validity domains" or "validity limits". The reduction of NM to SR in virtue of (i) demonstrates that NM remains valid in the domain in which $(v/c)^2 \ll 1$.

> The boundary of D [the validity domain of a theory] is reached when p [a characteristic dimensionless parameter] is no longer negligible. It is therefore not sharp but is given as a *known* error estimate (an order of magnitude estimate): it is the error one makes by applying that theory (rather than the finer one). The classic example is D (Newtonian Mechanics) which is given by $p = (v/c)^2$. (Rohrlich (1988), pp. 301–2)

In an earlier article Rohrlich and Hardin (1983) speak as follows:

> [R]elativistic particle dynamics (special relativity) leads to Newtonian particle dynamics in the limit as terms of order $(v/c)^2$ and higher are dropped while terms of order v/c are kept. The strong inequality $(v/c)^2 \ll 1$ thus characterizes the validity limits of Newtonian mechanics. . . .
>
> A *validity limit* is thus equivalent to a specification of the error made by using the lower level [coarser] theory instead of the higher level [finer] theory. Any predictions by the lower level theory should be multiplied by a factor $1 \pm \delta$ where δ is an order of magnitude estimate of the error made. (Rohrlich and Hardin (1983), p. 607)

The use of the "$\ll$" notation and the characterization of $p = (v/c)^2$ as an order of magnitude estimate indicate that Rohrlich is using the language of asymptotic analysis and perturbation theory. It is worthwhile to give precise meanings to these terms. Given two functions $f(x)$ and $g(x)$, $f(x) \ll g(x)$ as $x \to x_0$ means that $\lim_{x \to x_0} f(x)/g(x) = 0$. To say that $f(x)$ is at most of order $g(x)$ as $x \to x_0$ ($f(x) = O[g(x)], x \to x_0$), means that $f(x)/g(x)$ is bounded for x near x_0. In other words, $|f(x)/g(x)| < M$ for some constant M if x is sufficiently close to x_0 (See Bender and Orszag 1978).

Once one adopts this language to characterize the limiting relationships between theories, one must also pay attention to an important distinction in the theory of asymptotics. This is the difference between a regular and a singular perturbation series. A regular perturbation problem is one in which the exact solution for small but nonzero values of $|\epsilon|$ "smoothly approaches the unperturbed or zeroth-order solution as $\epsilon \to 0$". In the case of a singular perturbation problem "the exact solution for $\epsilon = 0$ is *fundamentally different in character* from the 'neighboring' solutions obtained in the limit $\epsilon \to 0$" (Bender and Orszag 1978, p. 324).

Two very simple examples can help illustrate the difference. Suppose we wanted to find the roots to the quadratic equation $x^2 + x - 6\epsilon = 0$, where ϵ is a small "perturbation" parameter. When $\epsilon = 0$ the equation has two "zero-order" ($O[\epsilon^0]$) roots: $x = 0, -1$. It is possible to solve the perturbed problem to different orders of ϵ, e.g. $O[\epsilon^1]$ and $O[\epsilon^2]$, by expanding the equation in a power series about the zero-order roots.[4] It will turn out that the series will have a finite radius of convergence and the solutions will smoothly approach the zero-order roots when $\epsilon \to 0$.

On the other hand, the quadratic equation $\epsilon x^2 + x - 1 = 0$ yields a singular perturbation problem. In the limit where $\epsilon = 0$, there is clearly only one zero-order root; namely, $x = 1$. But, the perturbed problem clearly has two roots. The equation suffers a reduction in order upon setting the parameter ϵ equal to zero. It is possible in this case and in many others to solve such problems, but the important thing for us to note is that there remains a fundamental difference between *the limiting behavior* as $\epsilon \to 0$ and *the behavior in the limit* where ϵ is identically equal to zero.

Now, the theory of elementary asymptotics shows that the fundamental formula appearing in the Lorentz transformations of SR, $\sqrt{1 - v^2/c^2}$, can be expanded in a Taylor series as $1 - 1/2(v/c)^2 - 1/8(v/c)^4 - 1/16(v/c)^6 - \cdots$. Thus, the limit $(v/c)^2 \to 0$ is analytic and the perturbation expansion is regular and not singular. I believe that this is extremely important. In effect it amounts to the claim that (at least some) quantities or formulae of

SR can be written as Newtonian or classical quantities plus an expansion of corrections in powers of $(v/c)^2$.

As a result, I think that there is warrant to Rohrlich's claim that the mathematical framework of SR implies the mathematical framework of NM when the validity domain of relatively theory is restricted to that of NM; i.e., when $(v/c)^2 \ll 1$. SR can be understood as a regular perturbation of NM with $p = (v/c)^2$ as the perturbation parameter.

Rohrlich is certainly aware that limiting relations between theories can be very subtle. In fact, he points out that in the $(v/c)^2 \to 0$ limit of the mathematical framework of SR the invariance properties of the framework will actually change from the Lorentz symmetries to Galilean symmetries. "[T]he limiting process is highly nontrivial and must be carried out very carefully: the symmetry reduction may be the result of group contraction, and the limit can only be carried out in suitable group representations" (Rohrlich (1988), p. 304).[5] However, I do not believe that he pays sufficient attention to the differences between regular and singular perturbation problems. In his list of examples where the relation: $(p \to 0)$ lim M(fine) = M(coarse) holds, he includes in addition to the SR/NM case, the relation between wave optics and geometrical optics as well as that between QM and CM. In contrast to the SR/NM case, these perturbation problems are singular: As we will see neither the solutions of wave optical equations nor of quantum mechanical equations can be expressed in terms of those of their corresponding coarser theories plus corrections in powers of some parameter p.

Now, Rohrlich is certainly cognizant of the fact that some kind of discontinuity exists between, e.g., wave optics and geometrical or ray optics, but he holds that this discontinuity is to be found wholly at the level of interpretation or semantics:[6]

> [D]espite the *continuous* change from one [mathematical framework] M to the other in the $p \to 0$ limit, the interpretation is discontinuous in the limit. This discontinuity is due to the *cognitive emergence* of a qualititatively new and different description. The equations obtained in the limit have a new and different interpretation (semantics) (Rohrlich (1988), p. 307).

I want to claim, however, that the *singular* nature of the *mathematical* limit between wave and geometrical optics is responsible for the "emergence" of a different description. Furthermore, as we will see, the presence of a third *theory* of the asymptotic domain between these two theories shows that the abruptness or discontinuity at the semantic level may in a certain sense be less pronounced than Rohrlich asserts.

Rohrlich's claim that the mathematical framework of CM can be "rigorously deduced" from that of QM in the limit where Planck's constant

$\hbar \to 0$ (remembering the earlier caveat, note 2, about this being shorthand for some dimensionless quantity with $\hbar$ in the numerator approaching zero) has been criticized by Hans Radder on different grounds. Radder rightly notes that without qualification this claim cannot be correct. He says that "the basic problem is that classical observables are mathematically represented by *functions* (on a phase space of generalized coordinates), while quantum observables are represented by *operators* (on a Hilbert space of state vectors). Since functions and operators are different mathematical entities, it is not possible, even when h is small, to derive an equation of functions from an operator equation" (Radder (1991), p. 219). It seems to me that part of this worry can be diffused by more carefully considering Rohrlich's claim that the mathematical framework of the coarser theory is "rigorously derived" from that of the finer theory. Perhaps, Radder understands this as meaning "can be validly deduced in, e.g., first order predicate calculus". But Rohrlich is quite explicit that his idea of a rigorous derivation involves a limiting procedure – a "rule of inference" not present in first order logic. One defense of Rohrlich is to take his "rule of inference" to be the possibility of finding a *regular* perturbation expansion for the formulae of the finer theory having first terms involving formulae from the coarser theory. If this possibility obtains (as in the SR/NM case), it does seem reasonable to speak of a *rigorous* derivation. In any event, the worries I have expressed about the singular nature of the limiting relationship between QM and CM and between wave and geometrical optics remain, and it is to a detailed investigation of these relationships that we now turn.

2.

The two cases, geometrical optics/wave optics and QM/CM are very similar. Both limiting relationships fall under the generic heading of shortwave asymptotics. Classical wave optics was preceded historically by geometrical optics just as QM was preceded by CM. As the geometrical optics/wave optics case may be slightly more intuitive, I will first discuss some features of this relationship and then use the results to assist in drawing conclusions about the QM/CM case.

The shortwave limit $\lambda = 0$ of wave optics is the theory of geometrical optics. The concept of a wave in this limit makes no sense and the basic entities of the theory are light *rays*. One of the most important aspects of ray optics is the fact that rays considered individually, are unimportant. Instead, the theory primarily refers to entire families of rays. Consider, for example, the common phenomenon of the focusing of a lens. The very concept of a focal point makes no sense if one considers a single ray.

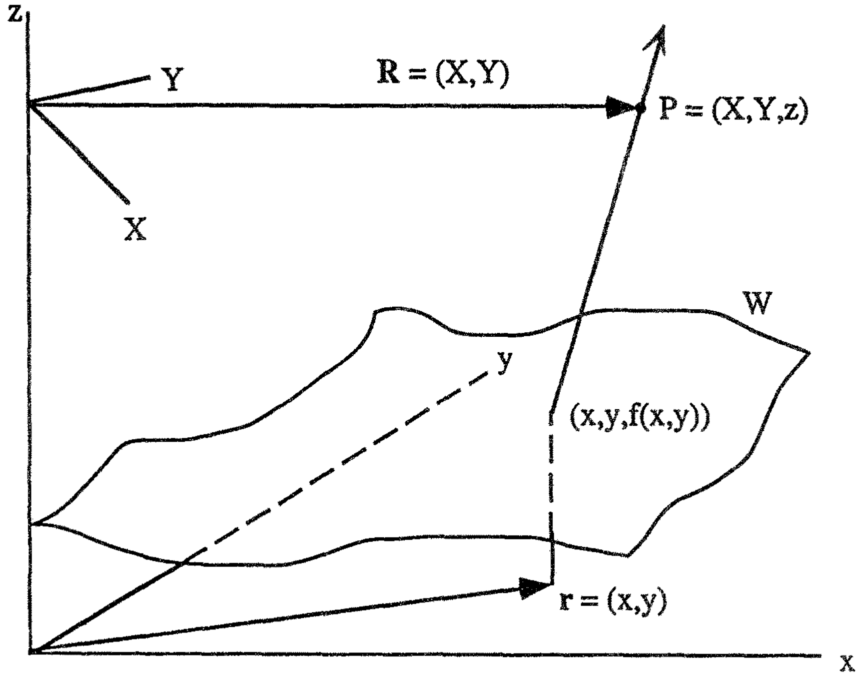

Fig. 1.

If geometrical optics is a *limiting* case of electromagnetism or wave optics, then we can ask what waves will look like in the limit as $\lambda \to 0$. They ought somehow to be constructed out of (or at least to correspond to constructions out of) the objects of geometrical optics – namely, families of rays. By studying the form of these constructions we will see that the limiting relationship is in fact a singular perturbation problem. Furthermore, we will discover reasons to doubt that any kind of reduction between the two theories is possible.

In the shortwave limit where $\lambda = 0$ or $k = \infty$ ($k = \lambda/2\pi$, the wavenumber), rays, not waves are the carriers of energy. Rays propagate as normals to "geometrical" wavefronts. (Think of a point source of light. The wavefronts will be concentric spheres with the source as their common center. Light rays will be those straight lines through the center and, of course, normal to each sphere's surface – the wavefront.) Consider Figure 1. (Berry (1981), pp. 518ff.)

We have an initial wave front W described by its deviation from the plane $z = 0$ by a height function $f(x, y)$. We want to construct, in terms of rays, an approximation to a wave at a point say P, which is a superposition of contributions from rays that pass through P. This is natural in that, as we have noted, when k is large the energy is carried along the rays. The rays emanating from the wavefront W are defined in terms of the gradient of the so-called optical distance function ϕ: Introduce a second coordinate system X, Y (with the same z-axis as the x, y system) to label the point P: $P = (X, Y, z)$. The optical distance function is:

$$\phi(x, y; X, Y; z) = [(z - f(x, y))^2 + (X - x)^2 + (Y - y)^2]^{1/2}.$$

In other words, it is just the euclidean distance between the point $(x, y, f(x, y))$ and the point $P = (X, Y, z)$. This assumes that the ray propagates in a homogeneous medium without refraction or reflection from W to P. If we imagine that W deviates only *gently* from the plane $z = 0$, we can use the paraxial approximation to ϕ. (Roughly since the inclination of the ray from the z-axis will be small, sines of angles may be replaced by the angles themselves.) This yields, letting $\mathbf{r} = (x, y)$ and $\mathbf{R} = (X, Y)$:

$$\phi(\mathbf{r}, \mathbf{R}, z) = z - f(\mathbf{r}) + \frac{|\mathbf{R} - \mathbf{r}|}{2z}$$

It is useful to think of the points $(\mathbf{R}, z)$, i.e. P, as control parameters or control variables – for instance, different places we might perform observations. Then the points $\mathbf{r}$ would label different possible starting points for paths to the point P. The different $\mathbf{r}$'s are called state variables. Fermat's principle tells us that the rays, for a fixed value of the control parameter $\mathbf{R}$ are those paths through P for which the optical distance function is an extremum:

$$\frac{\partial \phi}{\partial \mathbf{r}} = 0.$$

This gives

$$0 = \nabla_{\mathbf{r}}(\phi(\mathbf{r}, \mathbf{R}, z)) = -\nabla_{\mathbf{r}} f(\mathbf{r}) - \frac{\mathbf{R} - \mathbf{r}}{z},$$

Hence,

$$\nabla_{\mathbf{r}} f(\mathbf{r}) = \frac{\mathbf{r} - \mathbf{R}}{z}$$

and the rays are the solutions to this equation. As Figure 2 indicates there will generally be more than one ray through $P = (\mathbf{R}, z)$ which we label $\mathbf{r}^{\mu}(\mathbf{R}, z), \mu = 1, 2, \ldots$.

We are interested in constructing or associating a wave $\psi(\mathbf{R}, z)$ with this family of rays by allowing different rays through the same point P

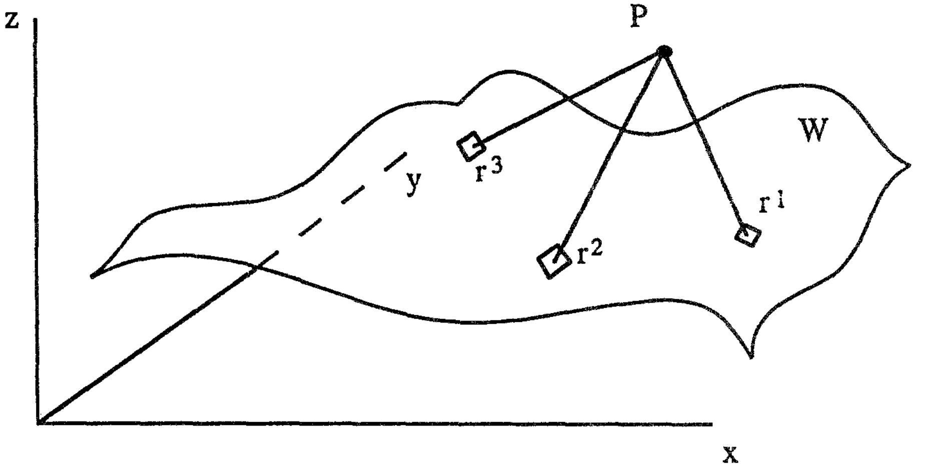

Fig. 2.

to "interfere" with one another. So we must use some kind of principle of superposition. Therefore, we need to know the contributions in both amplitude and phase of each ray $\mathbf{r}^\mu(\mathbf{R}, z)$. The phase of the μth ray is just 2π times the optical distance ϕ from the wavefront W to $(\mathbf{R}, z)$ in units of wavelength. That is, the phase is: $k\phi_\mu(\mathbf{R}, z)$ $(k = 2\pi/\lambda)$. The intensity (= $|\text{amplitude}|^2$) is given by considering the energy flux through an area dA(**R**) of a tube of rays around the μth ray when there is unit flux through dA(**r**) (see Figure 3). In other words, the amplitude is proportional to $|\mathrm{dA}(\mathbf{r}^\mu)/\mathrm{dA}(\mathbf{R})|^{1/2}$.

In fact, it will be given by the square root of the Jacobian determinant of the mapping from d$\mathbf{r}^\mu$ to d**R** given by the equation:

(i) $$\det \frac{\partial \mathbf{r}^\mu}{\partial \mathbf{R}} (\mathbf{R}, z).$$

So, the wave $\psi(\mathbf{R}, z)$ is, in shortwave approximation, given by the interfering ray sum:

(ii) $$\psi(\mathbf{R}, z) \approx \sum_\mu \left| \det \left\{ \frac{\partial \mathbf{r}^\mu}{\partial \mathbf{R}} (\mathbf{R}, z) \right\} \right|^{1/2} \mathrm{e}^{ik\phi_\mu(\mathbf{R}, z)}$$

It is possible to learn a lot from this equation. In the first place, as promised, its expresses the fact that the $k \to \infty$ limit is nonanalytic or singular. This is because k appears in the exponential. Unless the optical distance function $\phi_\mu(\mathbf{R}, z) = 0$ when $k = \infty$, the approximation ceases to exist. Therefore, the interference effects do not go away in the limit as $k \to \infty$; the wave

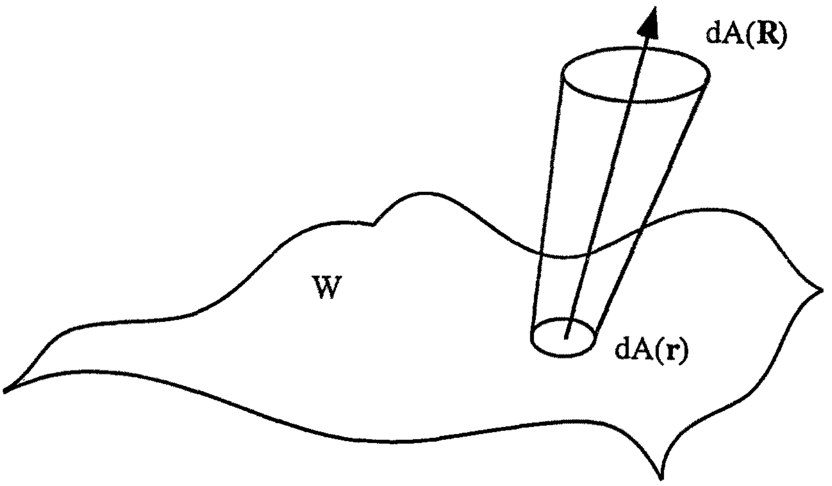

Fig. 3.

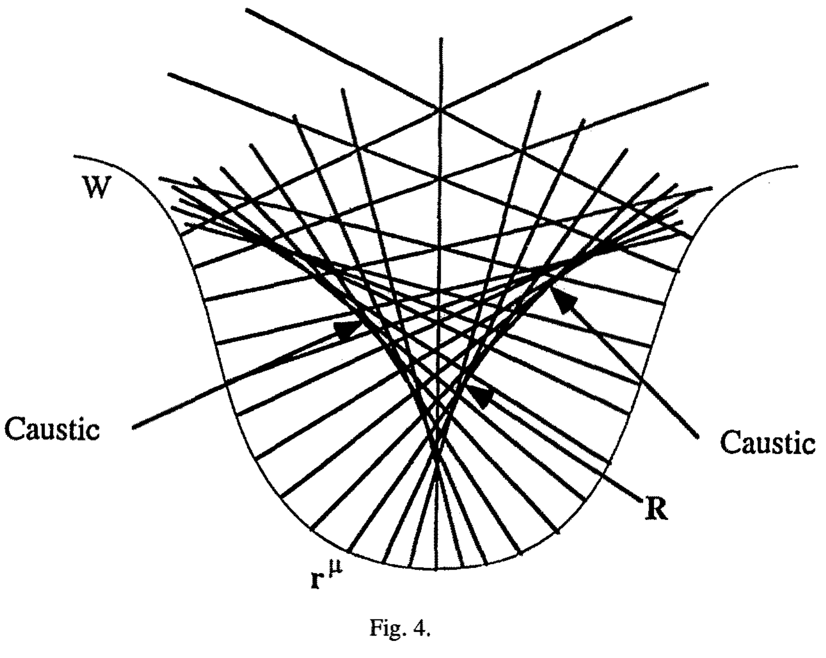

Fig. 4.

cannot be written in a series in which the leading term is purely a quantity describing rays and subsequent terms are corrections in powers of $1/k$.

Secondly, it may be the case that the point $P = (\mathbf{R}, z)$ lies on a focal surface or *caustic* of the family of rays. If this happens then the formula (ii) breaks down completely as the intensity becomes infinite. To illustrate the idea of a caustic or focal surface (a focal point is a special case), consider Figure 4. A geometrical wavefront W with normals (rays) is drawn. The caustic is the cusp shaped line which is the envelope of the family of rays. It serves to separate the region in which every point is the intersection of three rays (inside the cusp) from that where each point has only one ray passing through it (outside the cusp). Consider a ray $\mathbf{r}^{\mu}$ from W which touches a caustic at $\mathbf{R}$. The area $dA(\mathbf{R})$ of a ray tube around $\mathbf{r}^{\mu}$ shrinks to zero on the caustic. Therefore, the amplitude which, recall, is proportional to $|dA(\mathbf{r}^{\mu})/dA(\mathbf{R})|^{1/2}$ becomes infinite at the caustic. Caustics and focal points are primary objects of study in optics; but the interfering ray sum (ii) fails exactly at these places.

We began by asking what waves would look like in the limit as $\lambda \to 0$ or $k \to \infty$. The result (ii) while enlightening is not entirely satisfactory. It seems there is no way that the $\lambda \to 0$ limit between wave and geometrical optics can be regular. There is no way that λ or $1/k$ can be treated as a perturbation parameter. So, there is *prima facie* very good reason to disagree with Rohrlich when he places (as in the quote below) the wave optics/geometrical optics case as an example of reduction in the same class as the SR/NM case. He says:

> Another example is the reduction of geometrical optics, M(GO), to electromagnetic theory, M(EM). The characteristic parameter p is here the ratio of the wave length of light to the typical size of the objects considered. In the limit as p goes to zero, the fundamental equations of GO are derived from those of EM. The ray optics of GO is a limit of the wave optics of EM; at the same time, the explicit dependence on the wave length disappears from the equations of the theory. (Rohrlich (1988), p. 305)

The last clause is false. As we have seen, because of the singular nature of the limit, the wavelength cannot disappear from the equations as $\lambda \to 0$. Therefore, a straightforward reductive relation of the type Rohrlich envisions cannot be possible. There is still a limiting relationship between the theories. However, it is not of the proper sort to support a claim of reduction even in the attenuated sense in which all that is claimed is that the mathematical framework of the one theory can be deduced from the framework of the other via some limiting process. The asymptotic expansion is not regular.

There is another important and related reason for doubting that wave optics reduces to geometrical optics. This has to do with the fact, already noted, that the simple limiting relationship expressed by the interfering ray sum (ii) fails completely on caustics. If the control parameter $(\mathbf{R}, z)$ is varied accross a caustic, such as the cusp in Figure 4, the intensity becomes infinite. More sophisticated attempts to construct limiting approximations have been extremely successful in showing how this singularity is softened by diffraction patterns. In wave optics caustics do not exist since they are by definition singularities of ray families, and rays do not exist. But, in the limit as $\lambda \to 0$ it is possible to describe with remarkable accuracy the intensity of light on and near the geometrical caustics. The descriptions depend essentially on the nature of the caustic. *Thus, it appears that geometrical optics plays an essential role in the explanation of certain nonidealized physical phenomena.* (The pattern of light intensity in rainbows is a prime example.) Let me briefly describe the role geometrical optics has to play in these explanations. It is sometimes asserted that wave optics replaces or supersedes geometrical optics. However, the essential nature of the role of caustics in the explanations just mentioned is evidence that this assertion is simply wrong. We will see that there is really a third explanatory theory inhabiting the asymptotic domain between wave and geometrical optics.

Caustics are catastrophes. That is, they are classified by the mathematics of Rene Thom's catastrophe theory. Unlike focal points, some caustics are structurally stable in the sense that a slight perturbation of ϕ (due, e.g., to a small change in the shape of the wavefront) will not alter the basic form of the caustic. (A smooth change – a diffeomorphism – in ϕ results in a smooth deformation of the caustic.) The structurally stable caustics are the catastrophes. The transformability of one caustic into another by diffeomorphism results in a partition of caustics into equivalence classes. Two caustics belong to the same class just in case they can be so transformed. Each equivalence class is a catastrophe and can be represented by a certain polynomial that is linear in the control parameters and nonlinear in the state variables. These polynomials are called the normal forms of the catastrophes. For example, the normal form $\Phi(\mathbf{s}, \mathbf{C})$ of a simple fold catastrophe is, with $\mathbf{C}$ representing the control parameters and $\mathbf{s}$, the state variables, $\Phi_{\text{fold}}(\mathbf{s}, \mathbf{C}) = s^3/3 + Cs$. For the cusp catastrophe, $\Phi_{\text{cusp}}(\mathbf{s}, \mathbf{C}) = s^4/4 + C_2 s^2/2 + C_1 s$. This means that every optical distance function that yields a fold caustic can be transformed by diffeomorphism in the parameter C into $\Phi_{\text{fold}}(\mathbf{s}, \mathbf{C})$. Likewise, those ϕ's yielding cusps are transformable into $\Phi_{\text{cusp}}(\mathbf{s}, \mathbf{C})$. The same holds for higher order catastrophes such as the swallowtail, elliptical umbilic, and the hyperbolic umbilic. (See Poston and Stewart (1978) and Berry (1981).)

Armed with these results from catastrophe theory one can vastly improve on the interfering ray sum (ii) in two ways. First one replaces the function $\phi(\mathbf{s}, \mathbf{C})$ in the phase with the corresponding normal form $\Phi(\mathbf{s}, \mathbf{C})$. Second, since on a caustic the rays $s^{\mu}(\mathbf{C})$ coalesce, it is necessary to move to an integral representation where rather than a finite superposition of contributions coming from well-separated rays, one takes a continuous superposition integrating over all paths s. Away from the caustics, this inte-

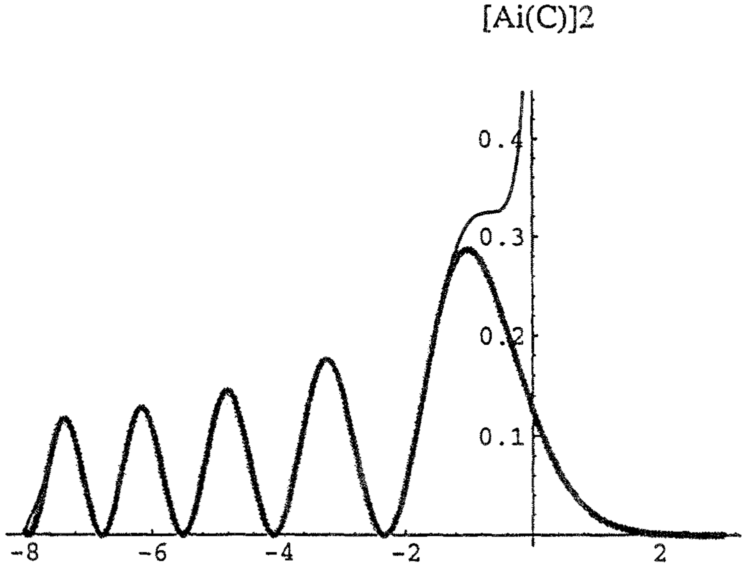

Fig. 5a. Square of the Airy function.

gral reduces to the interfering ray sum (ii).[7] But now, instead of diverging on the caustics as (ii) did, the integral representation gives one of a finite set of nondiverging "diffraction catastrophe" integrals depending upon the form of the caustic.

For the fold catastrophe, the integral is as follows. ($C = 0$ is the caustic.)

$$\text{(iii)} \qquad \psi_{\text{fold}}(C) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{i(s^3/3 + Cs)}\, \mathrm{d}s$$

Fermat's theorem, the ray condition, states that

$$\text{(iv)} \qquad \frac{\partial \Phi}{\partial s} = s^2 + C = 0$$

For $C < 0$ this means that there are two interfering rays $s^\mu = \pm\sqrt{-C}$. For $C > 0$, the roots of (iv) are complex and there are no real rays. The intensity $|\psi_{\text{fold}}(C)|^2$ gotten from (iii) is shown in Figure 5. It is the square of the Airy function $[\text{Ai}(C)]^2$. (See Airy (1838) and Berry (1981), p. 529.)

The wide grey line in Figure 5a is $|\psi|^2$ – the square of the Airy function. The thin dark line shows the intensity predicted by the interfering ray sum – diverging at the caustic $C = 0$. As is evident from the figure, the intensity peaks just on the bright side ($C < 0$) of the caustic and decays exponentially in the region where $C > 0$.

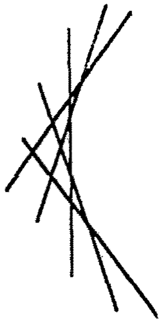

Fig. 5b. Rays forming a fold caustic.

The reason for going into this in as much detail as I have is to exhibit how an object or entity $\Phi(s, C)$ (the normal form of the caustic), which falls completely within the domain of geometrical optics, functions in a wave-theoretical explanation of the behavior of light near a caustic. The diffraction catastrophe integral (iii) is a hybrid formula. It is not constructable within orthodox wave optics. There is apparently no way to explain the behavior of such natural objects as the rainbow without combining in an essential way features of geometrical optics with those of wave optics. Therefore, it seems completely wrong to assert that the latter somehow reduces (in the sense of replaces or supersedes) the former.

It is sometimes asserted, correctly I believe, that since the coarser theory is really only an approximation, and therefore strictly speaking false, it cannot be explained by the finer theory. How can a false theory be explained? On the other hand, some have claimed that the finer theory can "explain away" the coarser theory (Sklar 1967, p. 112). Sklar, for example, means by this that the finer theory can explain why the coarser theory "met with such apparent success for such a long period of time and under such experimental scrutiny" (Sklar 1967, p. 112). Now Sklar makes this assertion about QM explaining away CM, but I think he might have said the same thing about wave and geometrical optics as well.

One way of understanding this is as follows. The finer theory can explain away the coarser by showing that the *numerical predictions* of the coarser theory are, within its domain of validity, essentially indistinguishable from those numerical predictions of the finer theory in that domain. In this sense the finer theory can explain why *numerically* the coarser theory worked as well as it did. But there is another, more robust, sense of "explains away". Consider the case of SR explaining away Newtonian mechanics. It seems that one wants to claim that NM is *correct* as an approximation in its *descriptions* of the phenomena within its domain of validity. Not only are its numerical predictions within the appropriate order of magnitude of those of SR, but its characterization of the phenomena *is sufficiently qualitatively similar* to that of SR as well.[8] This is, I believe a direct consequence of the mathematical fact that the SR/NM perturbation problem is regular. The limiting behavior as $(v/c)^2 \to 0$ is qualitatively similar to that in the classical limit when $(v/c)^2 = 0$.

On the other hand, in the wave/geometrical optics case, we have seen that the perturbation problem is singular. For $k \to \infty$ or $\lambda \to 0$ (as distinct from $k = \infty$ or $\lambda = 0$), the phenomena described exhibit wave properties such as diffraction. There is no way that geometrical optics can be correct as an approximation in its *description* of such phenomena, since such wave properties are completely foreign to geometrical optics. This does not mean that wave optics cannot "explain away" geometrical optics in the weaker sense of showing that its numerical predictions are within the appropriate order of magnitude of those of wave optics. However, it is another indication that when the finer theory is related to the coarser theory by a *singular* perturbation problem, it is inappropriate to claim that the latter reduces to the former in any sense of the term.

Trying to specify exactly what "is sufficiently qualitatively similar in its descriptions of phenomena" means is, of course, difficult. But I mean something more than Rohrlich when he speaks of descriptions that ("cognitively") emerge on the level of interpretation or semantics. I believe that the *mathematical form* of the equations characterizing the phenomena, and used in making predictions about their behaviors, is relevant as well. In particular, it seems reasonable to include as a criterion of qualitative similarity the possibility of relating the equations of the finer theory to those of the coarser by a regular perturbation problem. In such a case as we have seen, the equations of the coarser theory serve as leading terms in an analytic expansion.

We have seen that elements from two incompatible limiting theories can be consistently and coherently combined in formulas such as (ii) and (iii). This, together with the fact that such a combination is essential for explaining and understanding certain phenomena, leads me to believe that there is really a third distinct theory in this asymptotic domain between the two giants, wave and geometrical optics. Berry and Upstill note that this "*catastrophe optics* is unfamiliar and unorthodox, cutting across traditional categories within the subject [of optics]" (Berry and Upstill 1980, p. 259). Because of its explanatory efficacy and its ability to provide understanding of a wide class of optical phenomena (inexplicable on the purely wave or ray theories), I believe catastrophe optics warrants being considered a separate theory. So, we see that paying attention to the fact that the perturbation problem between wave and geometrical optics is singular has led to an entire realm of physics that might have remained obscure had we assumed the problem to be similar to the SR/NM limiting relationship.

3.

I now want to show how similar conclusions may be drawn about the so-called reduction of CM to QM. I will demonstrate the singular nature of the perturbation problem in this case, and argue that because of this, just as in the previous case, there is really a third theory, semiclassical mechanics, inhabiting the asymptotic domain between CM and QM. The arguments are quite similar to those given above for the optics case, though I will address an additional problem related to the distinction between integrable or regular and chaotic motion which appears in CM but which apparently is absent in QM. This is already an indication that the QM/CM perturbation problem is a singular one.

To begin, note that the most fruitful framework for the geometrical description of classical motion is provided by the phase space. This is a multidimensional euclidean space in which the complete state of a system is represented by a point.[9] As the system evolves according to Hamilton's equations of motion, the point representing the system, carves out a one-dimensional trajectory in the phase space. The analog of classical phase space in the optics case discussed in the last section, would be the space gotten by combining (taking the cartesian product of) the space of state variables $\mathbf{s}^{\mu}$ with the space of control variables $\mathbf{C}$. Figure 4 showed the cusp catastrophe in the two dimensional control space with coordinates $\mathbf{R} = (X, Y)$, z fixed. We saw there that the caustic organizes the multivaluedness of the family of rays – inside the cusp each point is the intersection of three separate rays, whereas outside there is only one ray through each point. In the combined space $\mathbf{s}^{\mu} \times \mathbf{C}$, this is represented by the foldings of a surface over the control space as in Figure 6 (Berry 1986, pp. 15–6.) If one projects the surface down onto the (X, Y) plane, one finds that within the cusp, each point (X, Y) receives three separate contributions (corresponding to the three distinct rays through (X, Y)), while outside the cusp it gets only one. The caustic itself is a singularity in this projection as the different folds coalesce for those values (X, Y) that form the cusp.

In classical mechanics, as I noted, the state of a system is represented by a point $(\mathbf{q}, \mathbf{p})$ in phase space. Here $\mathbf{q} = (q_1, \ldots, q_N)$ is the generalized coordinate for the system and $\mathbf{p} = (p_1, \ldots, p_N)$ is the corresponding generalized momentum. Thus, the classical state is completely specified by 2N numbers. On the other hand, in quantum mechanics, the system's

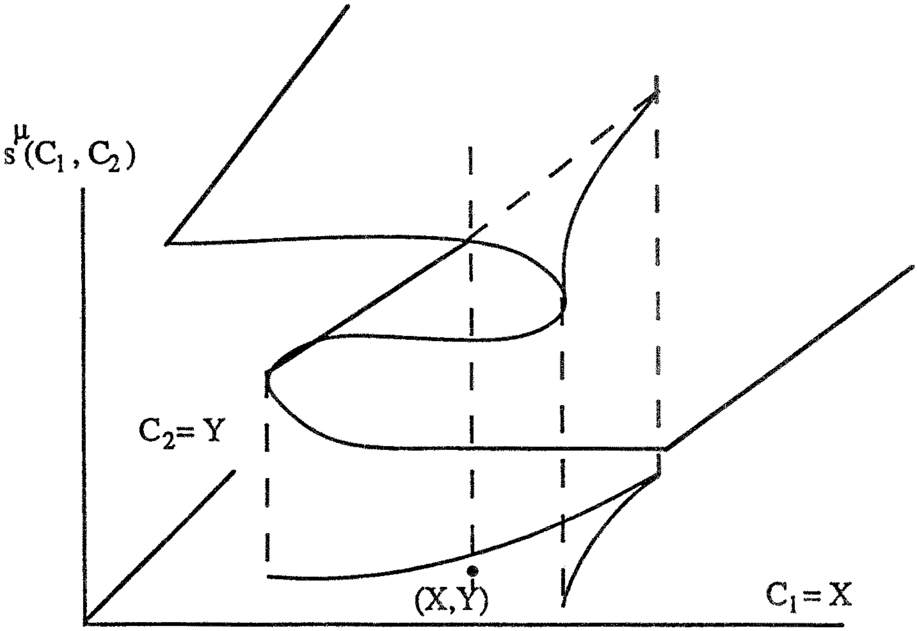

Fig. 6.

state is represented by a wavefunction $\psi(\mathbf{q})$ which is a function of only half the canonical coordinates ($\mathbf{q}$, $\mathbf{p}$). We will see shortly how this reduction in the number of coordinates is accomplished. The "objects" or entities of CM relevant for this discussion are families of trajectories in phase space evolving under Hamilton's equations of motion. These are analogous to the ray families of the preceding section. Clearly, the analog of the waves in wave optics are the quantum mechanical wavefunctions $\psi(\mathbf{q})$.

According to Rohrlich, just as in the previous case, the mathematical frameworks of the two theories are related by a limiting process

> $p \to 0$ lim M(QM) = M(CM) where p is (roughly speaking) the ratio of the size of a single quantum of some observable in QM to its size in CM. It is dimensionless but proportional to Planck's constant h. In the limit, the theory becomes independent of Planck's constant h. (Rohrlich (1988), p. 305)

As before, I will show that this asymptotic limit is singular, so that the last claim is once again false or misleading and hence, a straightforward claim of reduction as in the SR/NM case is not warranted.

We begin by asking the same question we did earlier. If CM is a limiting case of QM, then what will the wavefunctions look like in the limit when $\hbar \to 0$? We can expect to find a construction analogous to the interfering ray sum. That is, we are looking for an association between wavefunctions

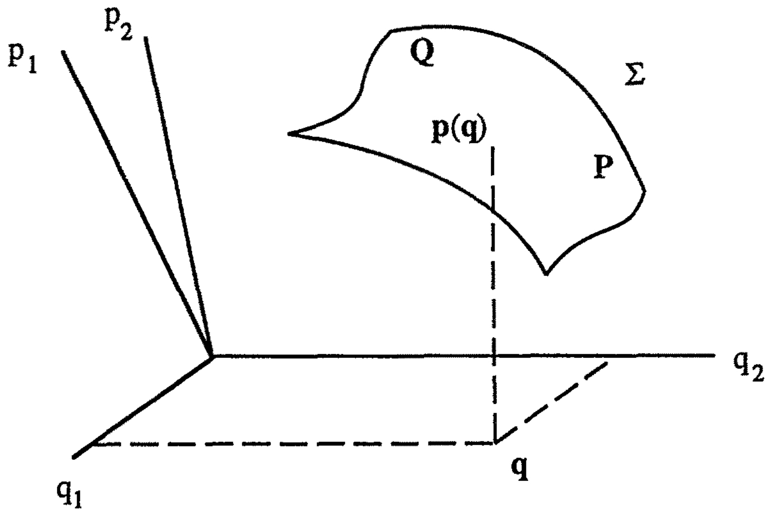

Fig. 7.

and families of phase space trajectories or, equivalently, surfaces in phase space.[10]

The association depends crucially on geometric properties of certain kinds of N-dimensional surfaces in the 2N-dimensional phase space – so-called Lagrangian surfaces.[11] We will first develop the connection in the abstract, and then try to provide it with a physical interpretation. Consider a local patch of a two dimensional Lagrangian surface Σ embedded in a four dimensional phase space as in Figure 7. It is possible to consider Σ as a function of the coordinates **q**; namely, **p**(**q**). *In so doing we treat the surface as a function of only half the canonical coordinates* (**q**, **p**), *just as quantum mechanical wavefunctions* $\psi(\mathbf{q})$. Next, introduce a new set of variables $\mathbf{Q} = (Q_1, Q_2)$ to label points on Σ. (These variables are analogous to the introduction of the coordinates X, Y used to locate the point **R** in the optical example of the last section. See Figure 1.) Corresponding to the new set of coordinates will be a set of conjugate momentum coordinates $\mathbf{P} = (P_1, P_2)$. The full set (**Q**, **P**) might, for example, be a set of angle/action variables. We can now think of (**Q**, **P**) as providing a different coordinatization of the entire phase space, which is connected to the original coordinatization (**q**, **p**) by a canonical transformation provided by a generating function S(**q**, **P**) – a function of both old and new canonical variables.[12] In other words,

(v) $\mathbf{p} = \nabla_{\mathbf{q}} S(\mathbf{q}, \mathbf{P})$ and $\mathbf{Q} = \nabla_{\mathbf{P}} S(\mathbf{q}, \mathbf{P})$ or schematically

$$(\mathbf{q}, \mathbf{p}) \leftarrow S(\mathbf{q}, \mathbf{P}) \rightarrow (\mathbf{Q}, \mathbf{P})$$

The Lagrangian surface Σ is the set of points in phase space of the form $(\mathbf{q}, \mathbf{p}) = (\mathbf{q}, \nabla_{\mathbf{q}} S(\mathbf{q}, \mathbf{P}))$. The generating function S(**q**, **P**) plays the role in mechanics of the optical distance function $\phi(\mathbf{s}, \mathbf{C})$ in geometrical optics.

Now the idea is to associate a (semiclassical) wavefunction $\psi(\mathbf{q})$ with this Lagrangian surface in such a way that the wave intensity, or rather the probability density, is proportional to the density of points in the classical *coordinate* space (in Figure 7 this is the (q_1, q_2) plane). This density is gotten by "projecting down" $(|d\mathbf{Q}/d\mathbf{q}|)$ from Σ onto q-space assuming points on Σ are uniformly distributed in $\mathbf{Q}$. Writing $\psi(\mathbf{q})$ schematically as

(vi) $\psi(\mathbf{q}) = a(\mathbf{q})e^{ib(\mathbf{q})}$

with amplitude a(**q**) and phase b(**q**), then this requirement on the probability density yields

(vii) $$a^2(\mathbf{q}) = K \left| \det \left\{ \frac{\partial^2 S(\mathbf{q}, \mathbf{P})}{\partial q_i \partial P_j} \right\} \right|; \text{ K constant.}$$

Here we have used the second equation in (v) in computing $|d\mathbf{Q}/d\mathbf{q}|$.

The phase b(**q**) of the wavefunction is gotten by appeal to the de Broglie relation $\overrightarrow{\mathbf{p}} = \hbar \overrightarrow{\mathbf{k}}$ which relates the momentum **p**(**q**) to the wave vector of a locally plane wave. This introduces Planck's constant and yields after some straightforward manipulation involving the relation, $\mathbf{p} = \nabla_{\mathbf{q}} S(\mathbf{q}, \mathbf{P})$, from (v) the phase:

(viii) $$b(\mathbf{q}) = \frac{S(\mathbf{q}, \mathbf{P})}{\hbar}.$$

Plugging (vii) and (viii) into (vi) finally yields the wavefunction

(ix) $$\psi(\mathbf{q}) = K \left| \det \left\{ \frac{\partial^2 S(\mathbf{q}, \mathbf{P})}{\partial q_i \partial P_j} \right\} \right|^{1/2} e^{i/\hbar S(\mathbf{q}, \mathbf{P})}$$

This, of course, has a form very similar to a single term in the interfering ray sum (ii). In particular, note the appearance of $\hbar$ in the denominator of the exponent. This shows that the limit $\hbar \rightarrow 0$ of (ix) is singular.

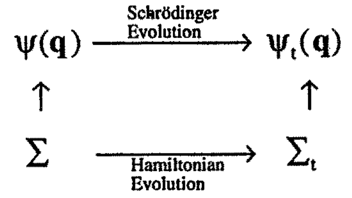

Fig. 8.

In the discussion of the interfering ray sum, no attention was paid to how the wave (ii) evolves over time. That is, no wave equation was considered. In the present context I want to remedy this situation because CM and QM are at base dynamical theories. The construction (ix) will be dynamically, and hence, physically significant if the diagram in Figure 8 is satisfied.

Let Σ be an initial classical surface evolving in time Δt into a surface Σ_t according to the classical Hamiltonian equations of motion. If $\psi(\mathbf{q})$ is the (semiclassical) wavefunction (ix) associated with Σ, then it will evolve in the same time interval Δt into the wavefuncion $\psi_t(\mathbf{q})$ according to the Schrödinger equation. The point is that in the limit as $\hbar \rightarrow 0$ (the semiclassical limit), the association between Lagrangian phase space surfaces and wavefunctions persists over time. Therefore, the wavefunction $\psi_t(\mathbf{q})$ can be determined from the time evolved surface Σ_t by the same construction (ix) used to construct $\psi(\mathbf{q})$ from Σ. In the semiclassical limit the association of a wavefunction with a phase space surface is time translation invariant. Basically, equation (ix) is an asymptotic solution of the Schrödinger equation to lowest order in $\hbar$.

Let me pause briefly here to elaborate on the significance of the construction we have just considered. We have shown that an equation intermediate between CM and QM, equation (ix), can be constructed in a sense "from the bottom up". That is, we can construct a new "semi-coarse" equation starting from structures (Lagrangian surfaces) present in the coarse theory CM. The fact that equation (ix) is *also* an asymptotic solution of the Schrödinger equation to lowest order in $\hbar$, as Figure 8 indicates, means that the construction *meshes* with the "semi-fine" equation one gets by beginning with the fully quantum mechanical Schrödinger equation, expanding it in powers of $\hbar$ and then dropping higher order terms. It is, of course, legitimate to drop higher order terms only if $\hbar$ can be considered small; that is, only if we are already in the semiclassical limit. The fact that these two procedures, one from the "top down" and the other from the "bottom up" agree, lends plausibility to the claim that the semiclassical equations provide something more than approximate methods for solving quantum problems. I will again take up this point in Section 5 below.

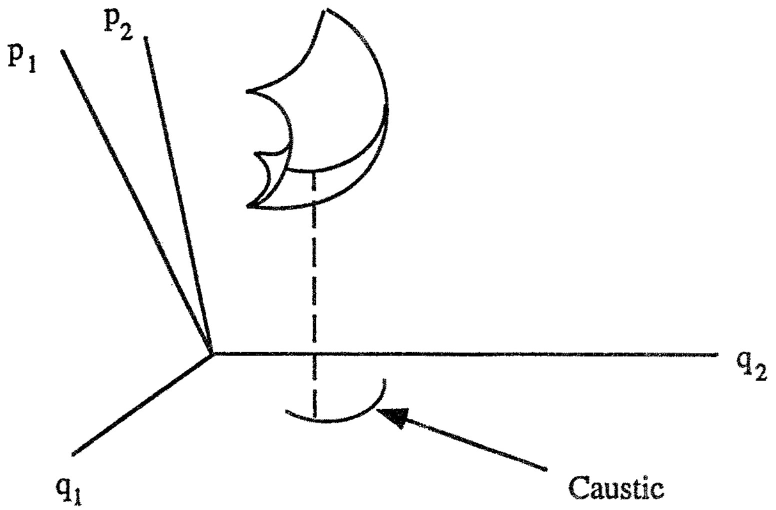

Fig. 9.

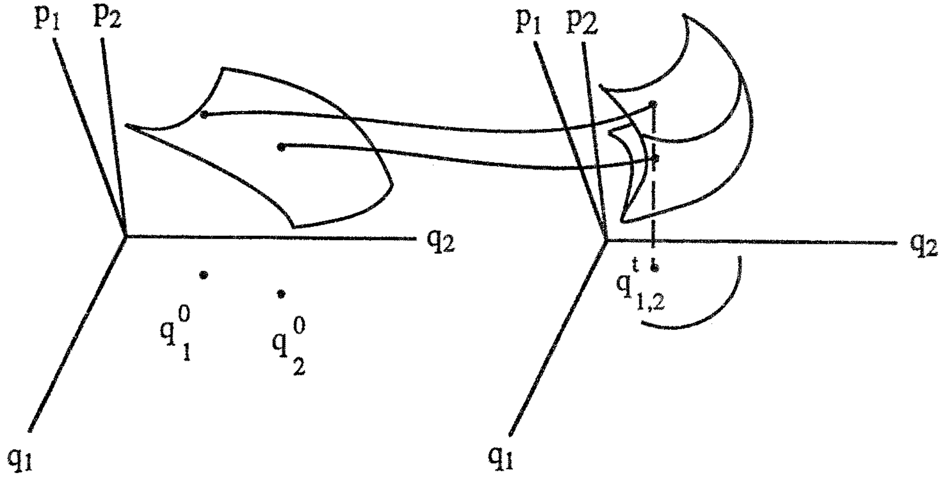

Fig. 10.

To return to the discussion, note that as Σ evolves over time it may develop folds as illustrated in Figure 9. Note two consequences of this evolution. First, the generating function S, and hence **p**, because of (v), becomes multivalued. This corresponds as in Figs. 4, 5b, and 6 to there being more than one ray through a given point. In the present case this means that two trajectories from different points on the initial surface Σ end up in the same time Δt at the same point in coordinate space. See Figure 10. Second, just as in the ray sum (ii), the determinant in the amplitude

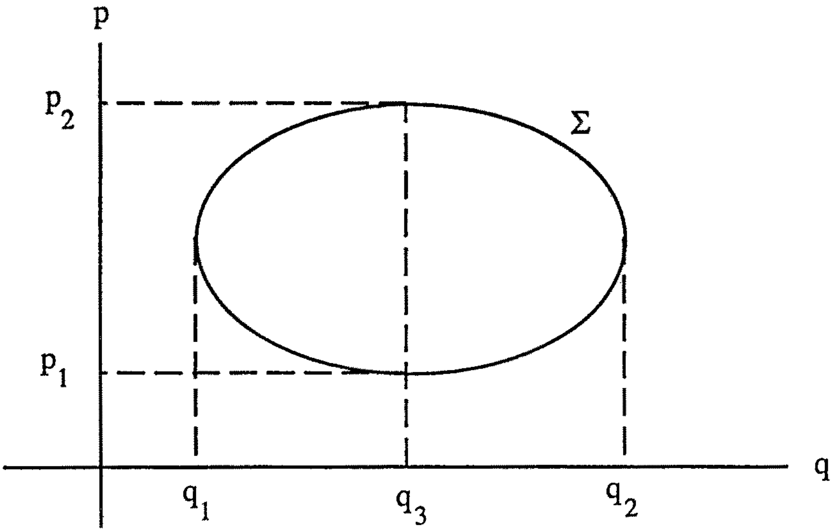

Fig. 11.

of $\psi(\mathbf{q})$ diverges to ∞ on a set of singular points of Σ – those points that project onto the (fold) caustic in q-space.

When folds in Σ develop and S becomes multivalued, the wavefunction *away from the caustic* may be represented by a *superposition* of terms of the form (ix) – one for each branch $p^{\mu} = \nabla_{\mathbf{q}}S(\mathbf{q}, \mathbf{P})$, a result very similar to the interfering ray sum (ii). Just as in the optics example, the presence of caustics signals a failure of the construction we have been discussing. An ingenious way of dealing with this problem was developed by Maslov.[13] One notes that the locations of the caustics of a Lagrangian surface depends crucially on the representation being used. For example, consider the phase space portrait of a one degree of freedom oscillator shown in Figure 11.

The q-space caustics of Σ are clearly the points q_1 and q_2 where the projection $|dp(q)/dq|$ is infinite. On the other hand, if Σ is given a momentum space representation: $\Sigma = (q(p), p) = (\nabla_p S(p), p)$, then the projection $|dq(p)/dq|$ is obviously not infinite at either q_1 or q_2. There are caustics in this representation – p-space caustics – at p_1 and p_2, but the corresponding q-values $q(p_1)$ and $q(p_2)$ do not coincide with q_1 or q_2.[14] By a construction completely analogous to the coordinate space construction resulting in (ix) one can associate a momentum space wavefunction $\chi(\mathbf{p})$ with the surface Σ. Since this will be well-behaved near the coordinate space caustics, it is possible to define the coordinate space wavefunction at and near the q-caustics as the Fourier transform of this momentum space wavefunction $\chi(\mathbf{p})$ (evaluated using the method of stationary phase).

The form of the probability density $|\psi(\mathbf{q})|^2$ will depend crucially, just as in the optical case, on the nature of the caustic. That is, *it depends on the geometric features of the projection of a classical phase space surface. These are purely classical features.* Thus, we have constructed, once again, a theory which combines coherently and consistently, concepts from the two incompatible primary theories. This theory, semiclassical mechanics, belongs to the asymptotic domain between CM and QM.

The existence of such a theory shows that there can be no reduction of QM to CM even in the physicists' sense. There are genuine measurable phenomena that are not fully understandable in either purely classical or purely quantum mechanical terms.[15] They fall, so to speak, between the explanatory cracks. Their explanation and understanding is provided by semiclassical mechanics.

4.

Until now, the fact that CM and QM are *dynamical* theories has played only a small role in our discussion. The one place where questions about dynamical evolution did arise, however, was crucial. The semiclassical construction receives a large part of its physical significance from its time translation invariance as expressed in Figure 8: The semiclassical wavefunction constructed according to Maslov's method evolves under the Schrödinger equation, as the Lagrangian phase space surface upon which it is based evolves under Hamilton's equations. But, it is only for a *very restricted* class of Lagrangian surfaces that this invariance holds over long periods of time. In most cases the invariance of the construction breaks down relatively quickly. The surfaces for which the diagram of Figure 8 holds for large times are those that are invariant or unchanging under Hamiltonian evolution. These surfaces have the topology of N-dimensional tori or doughnuts in the 2N-dimensional phase space. Dynamical systems whose possible trajectories are confined to such invariant tori are known as *integrable* systems. They exhibit regular – that is, periodic or multiply periodic – motions. The simplest example is a pendulum in one dimension whose torus in two dimensional phase space is the elliptical curve shown in Figure 11. A trajectory beginning from some point on the torus remains on it forever.

The semiclassical mechanics of systems whose "underlying" classical motions are integrable has been around for a long time. This "theory" is essentially a refinement of the old quantum theory of Bohr, Sommerfeld et al. It is usually called the EBK theory, after Einstein, Brillouin, and Keller;

although, a more descriptive name is Torus Quantization (See Berry 1983 and Batterman 1991).

The discussion above, however, has allowed for the possibility of studying evolving or noninvariant Lagrangian surfaces. This suggests, *prima facie* anyway, that it might be possible to extend these well-established results to systems whose underlying classical motions are *not* regular and integrable, but instead are chaotic. There are, however, serious difficulties involved in carrying out this suggestion. These problems provide further evidence that a reduction relation between CM and QM is not forthcoming.

Classical chaos is a "long time" property of dynamical systems. Were one to observe a system for any *finite* period of time and notice apparently random and unpredictable behavior, it is impossible to infer with certainty that the system is genuinely chaotic. A necessary condition (though not a sufficient one, see Batterman 1993a) for a system to be fully chaotic is that it possess strongly statistical ergodic properties, the weakest of which is ergodicity. But even ergodicity is defined in terms of infinite times; that is, in the limit as $t \to \infty$.

We have already seen that reductive relations between theories may depend on aspects of taking limits. The question now is how the taking of the $t \to \infty$ limit relates to the $\hbar \to 0$, semiclassical limit. The answer is that it does not relate well and as a result, the connection between classical and quantum mechanics is in a certain sense doubly singular. This requires some explanation.

Nonintegrable classical Hamiltonian systems are characterized by trajectories that are free to wander throughout the entire 2N − 1 dimensional surface of constant energy in phase space. This is in stark contrast to the trajectories of integrable systems which are confined to N-dimensional invariant tori. As a surface evolves under a chaotic nonintegrable Hamiltonian, it will generally become more and more convoluted, in the sense that the projection of the surface onto coordinate space (as in Figure 10) will develop an increasing number of caustics (see Berry et al. 1979 and Batterman 1993b). As long as the caustics are sufficiently well separated in coordinate space, the semiclassical construction of a wavefunction can proceed using Maslov's method to deal with the singularities at the caustics. In two dimensions there is a simple geometric criterion for determining when caustics are not sufficiently well separated. (See Berry et al. (1979) and Batterman (1993b).) Consider the curve Σ_0 in Figure 12a. Suppose it evolves so that after Δt the distinct points labeled 1–3 under classical evolution have the same q-value as in Figure 12b (Berry et al. 1979, p. 37).

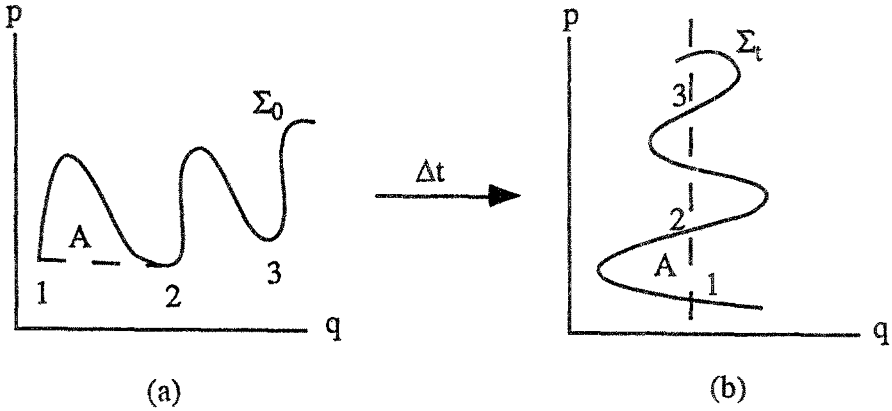

Fig. 12.

The constructed wavefunction $\psi_t(q)$ expressed as a sum of contributions of the form (ix) one for each point above q in Figure 12b, will be invalid if the phase space area A enclosed between the line connecting the preimages of the contributing points and the curve Σ_0 is of order $\hbar$. Because of the measure preserving nature of the Hamiltonian evolution this is equivalent to the claim that for contributing points sufficiently near q-space caustics, the simple semiclassical construction will fail. As long as these failures are isolated, they are exactly the catastrophes that can be dealt with by moving to an integral representation of the diffraction catastrophe integral type as in (iii) of Section 2. But, in the $t \to \infty$ limit, such catastrophes are no longer isolated. The Lagrangian surface becomes so convoluted that there will be ubiquitous clustering of caustics on scales smaller than order $\hbar$. In that case, the method of diffraction catastrophe integrals itself will break down.

The idea of a clash between the semiclassical and infinite time limits is really quite simple. For *fixed* time t after the initial time t_0, it will always be possible to construct a semiclassical wavefunction associated with the evolved classical surface, by letting $\hbar \to 0$. This is because for sufficiently small values of $\hbar$, the contributing points will be "far enough" from the caustic so that the area (as in Figure 12b) will be greater than order $\hbar$. On the other hand, if we let $t \to \infty$ first, this construction will fail because caustics will cluster on too fine a scale. Since classical chaos is a $t \to \infty$ property, it is apparently not recoverable in the $\hbar \to 0$ limit of quantum mechanics, and so we have yet another reason to doubt that there exists a genuine reductive relationship between the two theories provided by taking limits.

Now the fact that the semiclassical construction that we have discussed apparently breaks down when the underlying classical motion is chaotic might *prima facie* seem to count against the claim that semiclassical mechanics is a valid third theory "in between" CM and QM. Instead the singular behavior expressed in the failure of the two limits to commute is a further indication that there is interesting physics to be done in the asymptotic ($t \to \infty$, $\hbar \to 0$) domain. A similar attitude has recently been expressed by O'Connor et al. who say that

> it would be rather disheartening to be forced to accept that such a basic structure of chaotic classical mechanics as a homoclinic recurrence [a characteristic feature of unstable motion in a bounded domain] lies beyond what can reasonably be translated into quantum mechanics (via semiclassical mechanics). The consequences of such an *exceedingly* short time of validity would profoundly affect the utility of semiclassical methods. (O'Connor et al. 1992, p. 341)

Their paper is motivated by a desire to test the idea "that semiclassical mechanics must generally fail by the time classical structures are being formed on a scale small compared to h" (O'Connor et al. 1992, p. 342). The conclusion they draw is that contrary to popular opinion, "the development of classical structures small on the scale of Planck's constant [does not herald] the end of quantum-classical correspondence" (O'Connor et al. 1992, p. 355).

This quantum-classical correspondence finds its expression in the validity of semiclassical mechanics. O'Connor et al. purport to show, through detailed consideration of an (admittedly idealized) example, that despite the fact that "quantum mechanics smooths over classical fine structure ... classical fine structure *can* be used to construct quantum wavefunctions" (O'Connor et al. 1992, p 354).[16] Thus, current work on the semiclassical connections between QM and CM indicates that semiclassical constructions are valid (when properly formed) in the long time domain. This, to my mind, is a further indication of the legitimacy of treating the resulting theory as a distinct, genuine, explanatory theory in its own right, and not merely as a set of methods for providing approximate solutions to purely quantum mechanical problems.

5.

I have been arguing against the claim that QM reduces to CM and that wave optics reduces to geometrical optics in the limit as a small parameter ($\hbar$, λ respectively) tends to zero. That is, I have been arguing that a physicists' reduction or Nickles' "reduction$_2$" does not obtain between members of these pairs of theories. Consequently, it would seem that a philosophers' reduction, even in the attenuated sense used by Rohrlich, fails as well. The failure of the limiting reduction has, I believe, at least one consequence of great philosophical interest. This is the presence of an intermediate theory in the asymptotic domain between the members of the theory pairs – respectively, catastrophe optics and semiclassical mechanics.[17]

One question arises immediately: Do these intermediate theories posit an ontology distinct from the ontologies of the two established theories? That is, considering the QM/CM case, are there somehow "new" or ontologically distinct systems that are the subject matter of semiclassical mechanics? I think that the answer is "no". No new ontological level is being described. There are phenomena, such as the morphologies of certain wavefunctions, and the statistics of energy levels of certain systems, which have not received, and cannot receive adequate explanations on either the classical theory or the quantum theory. A proper understanding of these phenomena is provided by semiclassical mechanics; but the explanations do not make reference to any entities or structures not present in either CM or QM already. Thus, there is a new and different *explanatory level* which indicates the presence of a third theory; although, there is no new ontology associated with that third theory.

While semiclassical mechanics is itself a theory capable of explaining certain physical phenomena, it can also profitably be viewed as a kind of metatheory. Semiclassical mechanics is the theory which describes the connections or correspondences between CM and QM. It does so by explicitly showing how classical structures (e.g. Lagrangian phase space surfaces) *emerge* from QM as $\hbar \to 0$. I would claim that further study of limits of small parameters between theories, and singular perturbation problems will yield a more exact concept of emergence than typically appears in the rather murky literature on emergence. Semiclassical mechanics, in describing a quantum/classical correspondence in the "direction" from QM to CM, shows explicitly and precisely the sense in which classical concepts can emerge as the result of studying the limit of a small parameter $\hbar$.

In the other direction, in describing a correspondence from CM to QM, semiclassical mechanics also plays an extremely important role. It shows, contrary to popular opinion, that classical structures play an essential role in understanding and explaining quantum evolutions and structures. The role these classical (phase space) structures play is manifest only in the semiclassical limit as $\hbar \to 0$. Nevertheless, they must be important for a full understanding of quantum evolutions and the nature of what are (apparently) purely quantum mechanical structures such as wavefunctions. The issues about chaos, briefly discussed in the last section (see also Batterman (1993b) for more details) especially makes this clear. There are morphological differences in wavefunctions, the explanations for which make essential reference to the nature of "the underlying classical motions". In other words, whether the classical motion is integrable or chaotic is key to explaining the differences in forms of the wavefunctions. In a certain sense, classical structures play a deep explanatory role in QM. Classical evolutions drive quantum evolutions. As I said, this role becomes manifest only in the semiclassical limit: For "relatively large" values of $\hbar$, the classical structures and the role they play are obscured or hidden as a result of the uncertainty relations. However, because of their emergence as $\hbar \rightarrow 0$, it seems reasonable to believe that a full understanding of quantum mechanics must ultimately refer to them.

Finally, let me reiterate a point briefly made in Section 3. It is crucial for establishing the claim that semiclassical mechanics should be deemed a theory in its own right. The talk of correspondences obtaining in the two "directions" from QM to CM and from CM to QM as mediated by semiclassical mechanics is to be understood in the following sense. If we begin with QM and solve the Schrödinger equation by expanding it in powers of $\hbar$, we arrive at a very useful approximation which has historically sometimes been given the honorary title of "theory" – namely, WKB theory, the ancestor of modern semiclassical mechanics. However, most physicists, I believe, have taken it to be merely a useful tool and do not regard it as having any kind of explanatory autonomy. The view I have been advocating is that it does have this further status. The reason is that, as I have tried to show, one can build up in the direction from CM to QM to the "theory" in the manner of the construction outlined in Section 3 of the paper. (The same, of course, goes for constructing catastrophe optics from the coarser geometrical optics theory.) The fact that the two procedures arrive at the same result allows one to provide a partial interpretation for the intermediate theory in terms of the entities and structures of the coarser theory. In other words, the "top down" approximation gains legitimacy as an *explanatory* theory because one can interpret *its* success as a result of the agreement with the "bottom up" construction. The title "theory" should, therefore, no longer be considered merely honorary.

NOTES

* I wish to thank Roger Jones and Joe Mendola for valuable comments on this and related work. Discussions with Bill Wimsatt also helped me get clear about certain issues related to intertheoretic reductions. Of course, they are not responsible for any mistakes and misinterpretations that still remain.

[1] The plausibility of this claim depends to a large extent on what one takes to be the "central content" of these two theories. For example, if one considers the theories to be primarily

about spacetime structure, then the momentum and relative velocity equations (involving the v/c terms) really can be considered only marginally relevant for considerations of intertheoretic reduction; and the argument that follows may not be too convincing. I am indebted to Roger Jones for stressing this point to me. His general skepticism with the idea that there are particular mathematical equations which capture the central content of a given physical theory is well presented in Jones (1991).

[2] It is really better to think of the limit as $(v/c)^2 \to 0$ rather than $v \to 0$ because as $(v/c)^2$ is a dimensionless quantity the limit will not depend on the units used to measure the velocity. Similarly when we discuss the limit as $\hbar \to 0$, since $\hbar$ is not dimensionless, this is really shorthand for the limit in which some dimensionless quantity containing $\hbar$ in the numerator approaches zero.

[3] This is my locution, not Rohrlich's.

[4] See Tabor (1989), pp. 90–2 for more details.

[5] Rohrlich (1989) offers a very detailed discussion of the reduction$_2$ of relativistic theories to Newtonian gravitational theory.

[6] This is true at least for Rohrlich's articles mentioned already. In a more recent and very interesting article Rohrlich (1990) he does note the importance of singular limiting "discontinuous" relationships between the mathematical structures of certain theories.

[7] The integral is evaluated using the method of stationary phase: The only points giving positive contribution to the integral in the limit $k \to \infty$ will be those corresponding to geometrical rays. These are the stationary points. Because of the rapid oscillations all points near these stationary points will have contributions that cancel by destructive interference.

[8] See note 1.

[9] If the system has N degrees of freedom, the phase space will be 2N-dimensional.

[10] In Batterman (1993b) I offer an elementary account of this association. Here I will go into a bit more detail. The discussion closely follows that of Berry (1983), pp. 195f.

[11] For our purposes here we do not need to discuss the exact definition of Lagrangian surfaces. See Littlejohn (1992), or Arnold (1989) for details.

[12] See any text on CM for a discussion of canonical transformations and generating functions, e.g. Goldstein (1950) or Arnold (1989).

[13] See Maslov and Fedoriuk (1981) and also Littlejohn (1992) for an excellent discussion of many of the subtleties of these issues.

[14] In higher dimensions (more degrees of freedom) it may indeed be possible that coordinate space caustics coincide with momentum space caustics. But Maslov showed that it is always possible to find a canonical transformation to another mixed set of q_i's and p_j's, $i \neq j$, such that in that representation there are no q-space caustics (Littlejohn 1992, p. 26).

[15] An example of the kind of phenomenon I have in mind here is the fact that certain quantum systems, e.g. those whose "classical counterparts" are chaotic (see Batterman (1993b) for an attempt to explicate this notion of a classical counterpart), have eigenstates which when plotted as probability contours show what is called "scarring". These are regions where the probability density is much greater than one would expect from a "chaotic wavefunction". It turns out that the scars or regions of high probability density are concentrated on and near classical periodic orbits. Much work in semiclassical mechanics has been devoted to explaining the presence of these scarred wavefunctions. Obviously, I do not have the space to explain this here. See the excellent review by Heller (1991) for details.

[16] Context indicates that by "quantum wavefunction" they really intend what I have been calling semiclassical wavefunctions.

[17] An obvious question here is whether this is a general consequence of singular limiting relations between theory pairs. My suspicion is that it is, though I am presently unable to

provide an argument to that effect. Nevertheless, the following example may support the point. Singular limiting relations and methods of dealing with them (the renormalization group) are characteristic of the "new" theory of critical phenomena, describing universality in the behavior of systems undergoing phase transitions. See Bruce and Wallace (1989) for an argument that such investigations genuinely count as new physics.

REFERENCES

Airy, G. B.: 1838, 'On the Intensity of Light in the Neighborhood of a Caustic', *Transactions of the Cambridge Philosophical Society*, **6**, 379–403.

Arnold, V. I.: 1989, *Mathematical Methods of Classical Mechanics, Second Edition*, (trans.) K. Vogtmann and A. Weinstein, Springer-Verlag, New York.

Batterman, R.: 1991, 'Chaos, Quantization, and the Correspondence Principle', *Synthese* **89**, 189–227.

Batterman, R.: 1993a, 'Defining Chaos', *Philosophy of Science* **60**, 43–66.

Batterman, R.: 1993b, 'Quantum Chaos and Semiclassical Mechanics', *PSA 1992*, vol. 2, 50–65.

Bender, C. M. and Orszag, S. A.: 1978, *Advanced Mathematical Methods for Scientists and Engineers*, McGraw-Hill, New York.

Berry, M. V.: 1981, 'Singularities in Waves and Rays', in R. Balian, M. Kléman and J. P. Poirier (eds.), *Physics of Defects (Les Houches, Session XXXV, 1980)*, North-Holland, Amsterdam, pp. 453–543.

Berry, M. V.: 1983, 'Semiclassical Mechanics of Regular and Irregular Motion', in G. Iooss, R. H. G. Helleman, and R. Stora (eds.), *Chaotic Behavior of Deterministic Systems, (Les Houches Session XXXVI, 1981)*, North-Holland, Amsterdam, pp. 171–271.

Berry, M. V.: 1986, 'Twinkling Exponents in the Catastrophe Theory of Random Short Waves', in B. J. Uscinski (ed.), *Wave Propagation and Scattering*, Clarendon Press, Oxford, pp. 11–35.

Berry, M. V.: 1991, 'Some Quantum-to-Classical Asymptotics', in M. J. Giannoni, A. Voros, and J. Zinn-Justin (eds.), *Chaos and Quantum Physics (Les Houches Session LII, 1989)*, North-Holland, Amsterdam, pp. 253–303.

Berry, M. V., Balazs, N. L., Tabor, M. and Voros, A.: 1979, 'Quantum Maps', *Annals of Physics* **122**, 26–63.

Berry, M. V. and Upstill, C.: 1980, 'Catastrophe Optics: Morphologies of Caustics and their Diffraction Patterns', in E. Wolf (ed.), *Progress in Optics XVIII*, North-Holland, Amsterdam, pp. 257–346.

Bruce, A. and D. Wallace: 1989, 'Critical Point Phenomena: Universal Physics at Large Length Scales', in Paul Davies (ed.), Cambridge University Press, Cambridge, pp. 236–67.

Goldstein, H.: 1950, *Classical Mechanics*, Addison-Wesley, Reading, MA.

Heller, E. J.: 1991, 'Wavepacket Dynamics and Quantum Chaology', in G. Iooss, R. H. G. Helleman, and R. Stora (eds.), *Chaotic Behavior of Deterministic Systems, (Les Houches Session XXXVI, 1981)*, North-Holalnd, Amsterdam, pp. 547–663.

Jones, R.: 1991, 'Realism About What?', *Philosophy of Science* **58**, 185–202.

Littlejohn, R. G.: 1992, 'The Van Vleck Formula, Maslov Theory, and Phase Space Geometry', *Journal of Statistical Physics* **68**(1/2), 7–50.

Maslov, V. P. and Fedoriuk, M. V.: 1981, *Semi-classical Approximation in Quantum Mechanics*, (trans.) J. Niederle and J. Tolar, D. Reidel, Dordrecht.

Nickles, T.: 1973, 'Two Concepts of Intertheoretic Reduction', *The Journal of Philosophy* **LXX**(7), 181–201.

O'Connor, P., Tomsovic, S. and Heller, E. J.: 1992, 'Semiclassical Dynamics in the Strongly Chaotic Regime: Breaking the Log Time Barrier', *Physica D* **55**, 340–357.

Poston, T. and Stewart, I.: 1978, *Catastrophe Theory and its Applications*, Pitman, London.

Radder, H.: 1991, 'Heuristics and the Generalized Correspondence Principle', *British Journal for the Philosophy of Science* **42**, 195–226.

Rohrlich, F.: 1988, 'Pluralistic Ontology and Theory Reduction in the Physical Sciences', *British Journal for the Philosophy of Science* **39**, 295–312.

Rohrlich, F.: 1989, 'The Logic of Reduction: The Case of Gravitation', *Foundations of Physics* **19**, 1151–70.

Rohrlich, F.: 1990, 'There is Good Physics in Theory Reduction', *Foundations of Physics* **20**, 1399-412.

Rohrlich, F. and Hardin, L.: 1983, 'Established Theories', *Philosophy of Science* **50**, 603-17.

Sklar, L.: 1967, 'Types of Inter-theoretic Reduction', *British Journal for the Philosophy of Science* **18**, 109–24.

Tabor, M.: 1989, *Chaos and Integrability in Nonlinear Dynamics: An Introduction*, John Wiley and Sons, New York.

Department of Philosophy
350 University Hall
230 North Oval Mall
Columbus OH 43210-1365
U.S.A.