

Robust Supervenience and Emergence*

Alexander Rueger†‡

Department of Philosophy, University of Alberta

*

† For help with earlier versions of this paper I thank Bob Batterman, Alex Byrne, Martin Carrier, Mohan Matthen, Glenn Parsons, Dave Sharp, and two anonymous referees.

‡

ABSTRACT

Non-reductive physicalists have made a number of attempts to provide the relation of supervenience between levels of properties with enough bite to analyze interesting cases without at the same time losing the relation's acceptability for the physicalist. I criticize some of these proposals and suggest an alternative supplementation of the supervenience relation by imposing a requirement of *robustness* which is motivated by the notion of structural stability familiar from dynamical systems theory. Robust supervenience, I argue, captures what the non-reductive physicalist wants from supervenience; most importantly, it provides a natural background for reconstructing the notion of (diachronic) property emergence in a way acceptable to physicalists.

1. Introduction

Much of the discussion of supervenience and emergence has taken place in the context of the mind-body problem. My concern in this paper, however, is not with the relation of ‘the mental’ and ‘the physical’ but rather with supervenience and emergence relations between properties that the physical sciences deal with. In order to make contact with other discussions, however, I shall start out with remarks about problems that have come up in analyzing the relation of mental and physical properties.

Non-reductive physicalists have often considered the notion of supervenience as essential for formulating a satisfactory version of their view. Supervenience was intended to imply not only a co-variation of properties of one domain (e.g., the physical) with those from another domain (e.g., the mental) but a ‘dependence’ of one domain on the other. Part of the reason why the original enthusiasm about supervenience waned was the discovery of problems the various versions of the concept had in capturing this sense of dependence of properties on each other. After some attempts to tinker with the definition of the supervenience relation itself, a frequently suggested remedy for the difficulty was to keep the relation unmodified but supplement it with a further relation or requirement, viz., that supervenience be ‘explainable’ in a way acceptable to physicalists, thus rendering explicit the sense of dependence involved in the ‘bare’ supervenience relation.¹

In this paper I’ll briefly discuss these suggestions, starting with a particular scenario —

the ‘wayward atom case’ — in which a bare supervenience relation has been seen as in need of supplementation. I suggest an addition to the supervenience relation that would, in this case, resolve the problem in an intuitively satisfying way; this is the requirement of ‘structural stability’ or ‘robustness’ for the relation. Robustness, I argue, is not merely an ad hoc solution to the wayward atom problem; the notion rather captures an important aspect of what the physicalist wants to achieve by adding the requirement of explainability to bare supervenience. Robustness and explainability are intimately connected. Violations of the stability requirement for supervenience relations will be tolerable for the physicalist only under conditions to be specified. Cases in which structural stability fails to apply to the supervenience relation (under the conditions mentioned) are significant and interesting, I argue, because they exemplify a physicalistically respectable version of *diachronic* or *evolutionary property emergence*. This notion of diachronic emergence (which is not in competition with the *synchronic* supervenience relation) can be shown to capture the main intuitive features of the concept, *viz.*, emergent properties will be novel and non-reducible.

2. Supervenience Supplemented

Those who hold ‘zombie worlds’ to be possible can use them to refute the claim that mental properties (M, for short) supervene on physical properties (P, for short). A zombie world is a world in which the same physical subvenient basis that we find in our world exists

without any supervenient mental properties. A radical variation in M between our world and the zombie world is associated with no variation whatsoever in P between the worlds — a straightforward violation of any sense of supervenience of M on P. If we relax the “no variation in P” condition slightly, such as to allow ‘small’ P-variations between our world and the zombie world, we can construct what is known as the wayward atom scenario. Suppose the only difference, at the P-level, between our world and another physically possible world consists in the position of a hydrogen atom somewhere in these worlds and that the other world again does not contain any mental properties (Kim 1987 [1993, 85], 1989 [1993, 277]). A tiny variation at the level of P-properties is now associated with an enormous variation at the level of M. This scenario, however — contrary to the previous one — is compatible with the *global supervenience* of M on P because

M globally supervenes on P iff two possible worlds which are indistinguishable with respect to P are also indistinguishable with respect to M.

That global supervenience holds in the wayward atom case was often seen as a proof that this notion of supervenience does not capture any reasonable (i.e., physicalistically respectable) sense of dependence of M on P. If there is global supervenience, and hence supposedly some kind of dependence, in the atom case, it is ‘almost’ the same as no supervenience, and hence no dependence at all, in the zombie case because the atom and zombie scenarios are ‘almost’ identical. (The problem so perceived does not only affect the notion of global supervenience

but can be illustrated as well for the relation of *strong supervenience* [cf. Paull and Sider 1992, 842].) The supervenience relation thus appeared too tolerant with respect to radical variations in the supervening properties. This insight led to a natural rescue attempt. Define supervenience not in terms of indiscernibility but in a more coarse-grained fashion so that the small differences between our world and the wayward atom world are not registered as differences at all; that is, define supervenience with respect to ‘similar’, rather than indistinguishable subvenient bases (Kim 1984 [1993, 89f.]):

M globally supervenes in the similarity sense on P iff two possible worlds that are similar with respect to P are similar with respect to M.

Although much will depend on how ‘similarity’ is specified, it is clear that the wayward atom scenario now will not qualify as a case of global supervenience (in the similarity sense) anymore. The over-sensitivity of the indiscernibility-based notion of supervenience to small variations in the base is eliminated.

The price to be paid for this success, however, is high. Adopting similarity-based supervenience implies that phenomena where a large variation at the supervenient level is associated with a small variation at the subvenient level in general are to be analyzed as cases in which the supervenience relation fails to hold. Thus, for instance, the qualitatively radical change of the properties of a substance in a phase transition (e.g., liquid to gaseous state) which depend on quantitatively small variations of a parameter (e.g., temperature variations

around the ‘critical temperature’ of a substance), would have to count as evidence that the properties of the substance in the gaseous state do not supervene (in the similarity sense) on a base of properties that it nearly shares with the liquid state. Such ‘critical phenomena’, of course, are ubiquitous in the physical sciences and a physicalist should hardly want to claim that they are manifestations of non-supervening properties (McLaughlin 1995, 35). Modifying supervenience in this way appears to generate — at least for the physicalist — too many cases of non-supervening properties within the domain of the physical sciences themselves.

On independent grounds, Paull and Sider have argued that the wayward atom scenario should not be analyzed as a case where the supervenience relation fails to imply a proper dependence between properties. The physicalist, they show, has to opt for one of the following alternatives:

(I) Assume that (a) M supervenes globally on P, (b) the M properties are ‘intrinsic’ properties of the physical objects, and (c) the wayward atom world is (nomologically) possible. From these premises a contradiction can be derived and the physicalist can either deny (a) that M globally supervenes on P, in which case there is no problem with dependence, or conclude (against b) that M are not intrinsic properties, or that (against c) the wayward atom world is, after all, not a (nomologically) possible world, in which case, again, the global supervenience relation cannot be blamed for failing to involve proper dependence.

(II) Alternatively, therefore, assume that (a) M supervenes globally on P, (b’) the M

are ‘extrinsic’ properties, and (c) the wayward atom world is (nomologically) possible. This leads to a consistent scenario where M does depend on P, albeit in a “rather bizarre”, that is, very sensitive, way (Paull and Sider 1992, 845).

Paull and Sider recommend that the physicalist adopt strategy (I) in this case because they feel that we have enough knowledge to make it at least probable that wayward atom worlds are (nomologically) impossible. Given the cases of sensitive dependence in nature, like phase transitions, however, the physicalist should probably not settle too easily for option (I). Suppose the possibility of the scenario cannot be excluded. In what ways can we then characterize what is “bizarre”, or unsatisfactory for the physicalist, about the dependence of M on P?

We could require, in addition to the bare supervenience of M on P, which may involve strange dependencies like in the wayward atom case, that the relation between M and P be ‘explainable’ in a way that is acceptable to the physicalist. Even though we now accept, say, that M supervenes on P in the wayward atom scenario, we require a further specification of the ‘ground’ of the supervenience relation, or an explanation for why the relation holds in this case, before we take the supervenience of M on P under these circumstances as an illustration of physicalism. Bare supervenience, according to this view, does not characterize physicalism because a monist, a substance dualist, and a defender of M-P parallelism can all agree on bare M-P supervenience. If a distinctively physicalist rendering of the relation between M and P is

to be found, it would have to include more than bare supervenience; the requirement of explainability is supposed to provide this supplement. The monist, the dualist and the parallelist will presumably differ in the way they respond to the request for grounding or explaining the supervenience of M on P (cf., e.g., Kim 1990 [1993, 156ff.], Kim 1997, 189f.).

Horgan has dubbed this composite relation of {supervenience and explainability} “superdupervenience” — “a kind of ontic determination which is itself... [physicalistically] kosher, and which thereby confers... [physicalistic] respectability on higher-order properties...” (Horgan 1993, 566) Thus, in order to decide whether the wayward atom scenario causes trouble for physicalism or not, we would have to find out whether the extremely sensitive dependence of M on P allows for a physicalist explanation — an explanation, roughly, like the ones we can give of the supervenience of macro-properties like liquidity on micro-properties of a substance. If such explanations are forthcoming or likely to be found, the physicalist can incorporate the M-P relation into his or her picture of the world. If there are no suitable explanations, the physicalist has to face a case of *emergent* properties. Whether or not the existence of emergent properties is to be taken as a refutation of physicalism is a contested issue and will be discussed further below. The important feature of this suggestion is that emergence does not coincide with a breakdown of the supervenience relation itself; rather, emergence is a failure of the {supervenience and explainability}-relation due to the second component being missing. This reconstruction of the notion of property

emergence arguably captures the sense in which some of the British emergentists used the term (Kim 1992).

As attractive as this proposal may initially appear, it has problems of its own. The most outstanding of these, in my view, is the difficulty in specifying what ‘explainability’ in this context amounts to. Are we, for instance, supposed to appeal to further laws of nature in order to explain the co-variation and dependence between M and P? Or are we asked to give a metaphysical grounding of the supervenience relation? The first option, explanation by empirical laws, may well result in nothing but a shift of the problem of grounding the M-P relation to explaining further relations (other laws) in physicalistically acceptable ways. The second option, a metaphysical grounding of the supervenience relation, is in danger of providing nothing but a mere restatement of the respective metaphysical positions that the {supervenience and explainability}-relation was supposed to help distinguish, *viz.*, the positions of the monist, the substance dualist, and the M-P parallelist.²

In addition to these difficulties explainability is often understood as an *epistemic* notion and it is not clear why such a notion should be needed to characterize an essentially *metaphysical* doctrine like physicalism. “[T]he thesis that a given domain supervenes on another”, wrote Kim, “is a metaphysical thesis about an objectively existing dependency relation between two domains; it says nothing about whether or how details of the dependency relation will become known so as to enable us to formulate explanations,

reductions, or definitions.” (Kim 1984 [1993, 76]) Analogously, the doctrine of emergence, if explicated in terms of a failure of explainability, seems to turn from a metaphysical position into an essentially epistemic doctrine. If we find this undesirable, we need to explicate explainability in a way which is not essentially epistemic and which fulfills the intended task of grounding supervenience in a way acceptable to the physicalist.

What is intuitively strange or bizarre about the wayward atom scenario, I suppose, is that the dependence relation between P and M seems to go against an expectation of robustness or stability that many people have with respect to relations in nature. Can this expectation be spelled out and justified so that (i) it does not get invoked as an ad hoc imposition, motivated solely by the desire to resolve the wayward atom problem, i.e., can robustness be shown to be connected to the physicalist search for explainable supervenience relations? (ii) Can robustness be explicated so that it does not rule out the possibility and actual existence of unstable phenomena in nature?

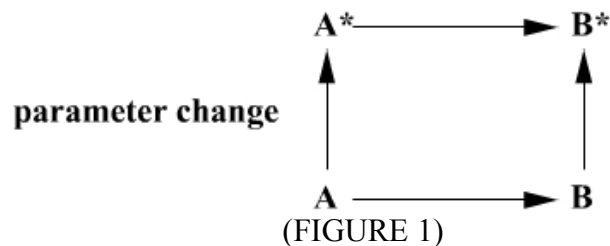
3. Supervenience and Structural Stability

I suggest to supplement the bare supervenience relation with the requirement that the relation can be explained, or ‘grounded’, as I prefer to say, and that the relation is grounded if (i) the relation is ‘structurally stable’, i.e., we require that a small change (a perturbation) of the subvenient base does not lead to a qualitative (‘radical’) change in the supervenient

properties, and (ii) the relation is such that in cases where condition (i) is violated, the qualitatively different behaviour or new properties in the system occur in a sufficiently stable way.³ The notions of structural stability and of its violations (‘bifurcations’) are formally defined in dynamical systems theory and will be explicated in more detail below and in Section 4. Since I shall employ these established notions I have to restrict my discussion from now on to subvenient and supervenient properties which can be represented in the framework of dynamical systems theory. The question whether mental properties fall into this category is left undecided.⁴ As long as this question is undecided, of course, the bearing of my discussion on the wayward atom problem is at best by way of analogy. But I think this scenario provides a valuable intuitive motivation for the robustness requirement, although the scenario by itself cannot carry the burden of justifying the condition of structural stability for supervenience relations.

I describe a physical system at a time by the variables used in dynamical systems theory: some set of ‘generalized coordinates’ — position and momentum — , which take on a sequence of values as time passes, a sequence represented as a trajectory in the system’s phase space portrait. In addition to these coordinates we need one or more ‘control parameters’ which specify features of the system, like friction, which are not assumed to be determined by the system’s internal dynamics. The combination of generalized coordinates and control parameters forms the subvenient base of the system. The phase space trajectories express the

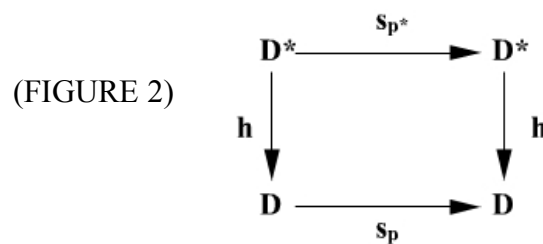
behaviour of the system — the collection of its supervenient properties. For instance, a behavioural property of a damped oscillator is that its motion gradually winds down and ends in a rest state while an undamped oscillator has the property of continuing its motion indefinitely. We could express things perhaps more precisely as follows: The property of having its motion wind down (call it ‘B’, a second-order property) can be ‘functionalized’ as the property of a system showing such and such long-term behaviour, given such and such input; we then find a structural (first-order) property which ‘realizes’ B, namely, the property A of being a system with its variables and parameters arranged in the way characteristic of a damped harmonic oscillator. The evolution of a system in this framework would be depicted as shown in Figure 1 (SV denotes the supervenience relation):



While the supervenience relation between properties of a system is a relation *at a given time* (synchronic), the condition of structural stability concerns the qualitative *change* in the behaviour of a dynamical system (the system’s phase space portrait) when the control parameter(s) change slightly over a period of time.⁵ Suppose that the system’s dynamics is characterized by an equation of motion with a control parameter p . If the phase space portrait

stays *qualitatively* the same under perturbations of the dynamics, i.e., small variations in the value of p ,⁶ the system is structurally stable. If the perturbation generates a qualitatively different portrait of trajectories, the system is structurally unstable. The notion of a *qualitative difference* between two sets of trajectories (phase space portraits) is spelled out in terms of the portraits being *topologically inequivalent*: one set cannot be transformed into the other by any smooth deformation of the trajectories.

In somewhat more technical terms⁷: A dynamical system Σ , considered as the transformations $s_p: D \rightarrow D$ on the system's phase space of initial conditions x at some initial time into solutions $s_p(x)$ of the dynamical equations at other times, is *topologically equivalent* to another system Σ^* with $s_{p^*}: D^* \rightarrow D^*$ if there exists a homeomorphism h (a one-to-one mapping continuous in both directions) of the phase space trajectories of the first system onto the trajectories of the second such that the diagram of Figure 2 commutes:



That is:

$$h [s_p(x)] = s_{p^*} [h(x)].$$

In other words: two systems are equivalent in this sense if the change from s_p to s_{p^*} , introduced by the variation of the control parameter p , can be compensated by a

transformation (h) of the coordinates. A system Σ is *structurally stable* if every system ‘close’ to Σ is topologically equivalent to Σ . (The notion of closeness has to be spelled out in whatever topology is imposed on the phase spaces.⁸)

As an illustration consider a linear dynamical system with parameters a, b, c, d , described by the differential equations:

$$dx/dt = ax + by$$

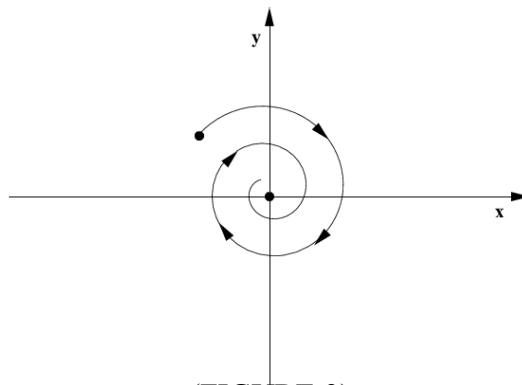
$$dy/dt = cx + dy$$

(Think of x as the position, y as the momentum variable.) The qualitative behaviour of this kind of system, characterized by the number and types of equilibrium states it possesses, will

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

depend on the eigenvalues of the matrix of coefficients

(i) Suppose the eigenvalues have *negative real* parts. The trajectories in this case all approach the origin of the phase space; the behaviour is that of a damped harmonic oscillator (Figure 3).



(FIGURE 3)

The necessary and sufficient conditions for this case (eigenvalues real and negative) are expressed as

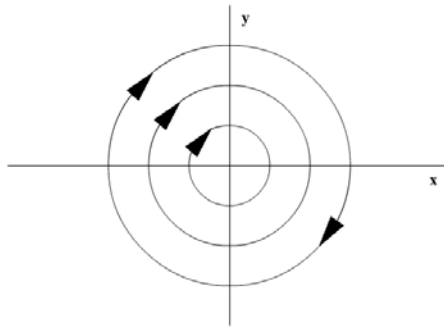
$$(a + d)^2 > 4(ad - bc)$$

$$(a + d) < 0$$

These inequalities are (comparatively) insensitive against replacing the coefficients by slightly different ones. That is, even if we perturb the system by changing a, b, c, d slightly, we still satisfy the inequalities which characterize qualitatively the system's behaviour. (To be sure,

there will be combinations of slightly varied parameter values which violate the inequalities; but for ‘almost all’ variations (in the measure-theoretic sense), they will be preserved.) The system is structurally stable against such perturbations — the original system and the perturbed system share essentially the same behaviour.

(ii) Suppose the eigenvalues of the matrix of coefficients are all *purely imaginary*. The phase space portrait is that familiar from the undamped harmonic oscillator (Figure 4).



(FIGURE 4)

The conditions for the parameters in this case are:

$$(a + d) = 0$$

$$(ad - bc) > 0$$

One of the conditions now being an *equality* rather than an inequality implies that a small

change in a and/or d will, in general, lead to a violation of the condition and thus to a qualitative change in the system's behaviour. The system is structurally unstable. If $(a+d) \neq 0$, the eigenvalues of the matrix will not be purely imaginary and will have a non-zero real part. This will turn the system into a system of kind (i) with trajectories spiraling into the origin of the phase space if the real part of the eigenvalues is negative, or into a system with trajectories diverging from the origin if the real part is positive. (Again, there will be certain combinations of changes in the parameters which preserve the condition; 'almost all' variations, however, will lead to a violation.)

This can also be expressed in terms of the unstable system (case (ii)) undergoing a *bifurcation* upon variation of its parameters. Some perturbations — those which are compatible with the conditions on the parameters — will leave the system unstable, others — in fact, most perturbations — will turn the system into a stable one. These two responses of the original unstable system to small disturbances are qualitatively different.

The intuition underlying the *first disjunct* in the definition of robust supervenience, once more, is this: If a set of properties B robustly supervenes on a set A, 'wiggling' some member(s) of the A set will not result in a *qualitatively* different B set, although *some* variation in B will be tolerated. The determinates of B properties may differ, i.e., their values may be different as a result of wiggling the A properties, but no new determinables arise, i.e., no new B properties are generated as the result of A-variations. A case like the wayward

atom scenario can therefore be analyzed as a violation of this first disjunct in the characterization of robust supervenience.⁹ The position of the wayward atom is the control parameter upon which the B properties sensitively depend; that is, whether there are B properties at all depends on this parameter taking on one particular value rather than a slightly different value. Note that it is the existence of B properties themselves, as qualitatively new compared to the situation without any B properties, which is due to the variation of the parameter. (The B domain does not depend on the position of the wayward atom in the sense in which the behaviour of a system depends on its initial conditions because variations of initial conditions do not generate a new (qualitatively different) dynamics for the system. The dynamics itself has to change, in dependence on the control parameter, in order for B properties to arise.¹⁰)

Why and in what sense can the robustness requirement provide a ‘grounding’ for bare supervenience? When the structural stability requirement is satisfied we have a supervenience relation which ensures that a certain range of slightly different base property distributions will be associated with distributions of the supervenient properties which are qualitatively similar to each other. Whatever ways there may be to explain a supervenience relation, pointing to the fact that a given relation is structurally stable clearly provides some ‘grounding’ for the relation. The fact that a supervenience relation is qualitatively stable against changes in circumstances is a relevant contribution to any explanation of why this relation holds and is

observed to hold. A system, in order to be realized in nature, cannot have the majority of its parameter values as bifurcation values; bifurcation points have to be exceptional in order for a system to survive the small perturbations that are always present in nature. The existence of robust systems requires less explanation than the existence of more delicate, unstable systems. Stability is always, as the practice of scientific explanation reveals, a contribution to explainability.¹¹

Nevertheless, of course, instabilities do occur and the robustness requirement as presented so far (its first disjunct) cannot be a *necessary* condition for grounding the supervenience relation in a physicalistically acceptable way. The second disjunct of the characterization of robust supervenience, therefore, deals with violations of structural stability.

We have to distinguish two ways in which the notion of robustness can be applied. As presented so far, structural stability of a system is defined for *individual* systems where a system is identified by a map on the system's phase space, $s_p: D \rightarrow D$, from initial conditions to solutions of the dynamical equations. Such maps contain one or more parameters p which, when considering an *individual* system, are taken to be fixed at certain values. Small perturbations can take the form of variations of those parameters which turn the original system into a different system; if the resulting system is topologically inequivalent to the original one under the perturbations considered, the original system is structurally unstable. Now consider, instead of individual systems, *families* of systems (or maps), where a family is

generated from an individual map s_p by adding perturbations to s_p which depend on further parameters α and letting these parameters run through their admissible values. We write s_α for the family to indicate the dependence of the class of systems on the parameters α . The harmonic oscillator with damping we considered above can be described by a family of maps with each individual oscillator corresponding to a definite value of the damping parameter.

Thus, a family or ‘unfolding’ of a system s_p consists of a set of systems which can be generated as possible perturbations from s_p . If s_p is given as a one-parameter equation $f(x; p) = 0$, an unfolding of s_p is the r -parameter family $U(x; p; \alpha_1, \dots, \alpha_r)$ with $U(x; p; 0, \dots, 0) = f(x; p)$.

Although the number and form of the perturbations considered in U is in general arbitrary, a *universal unfolding* can be defined as that unfolding of s_p that includes *all* possible (small) perturbations of $f(x; p)$ and uses the *minimum* number of parameters $\alpha_1, \dots, \alpha_r$. ‘All possible perturbations’ means all those perturbations which change the properties of $f(x; p)$ qualitatively. Although universal unfoldings are unique only up to coordinate changes, they can be classified into types of increasing complexity, measured by the minimum number of parameters required (for $r \leq 5$). Knowledge of such a type implies knowledge of the qualitative features of *all* the particular individual systems that get ‘unfolded’ in that type.¹²

In analogy to the case of individual systems we can now define a notion of *topological equivalence* for families of systems¹³: Two families s_α and g_α are topologically equivalent if there exists (i) a homeomorphism e mapping the parameter space of the first family onto the

parameter space of the second, and (ii) a *family* of homeomorphisms h_α , depending continuously on the parameter α , which map, for each value of α , the phase space trajectories of s_α onto the phase space trajectories of $g_{e(\alpha)}$ (where $e(\alpha)$ is the transformation of α under the homeomorphism e). A family of systems or maps s_α is *structurally stable as a family* if s_α is topologically equivalent to all families of systems or maps ‘close’ to s_α .

Intuitively, a family of systems is stable in this sense if it behaves qualitatively like nearby families. This extended notion of robustness now allows us to say that even though a given individual system may be structurally unstable (under a specified perturbation), the system could nevertheless belong to a stable family of systems. If it does, then the bifurcation or instability in the system occurs in a structurally stable way: all systems in the family exhibit qualitatively the same bifurcation behaviour. (This feature would be relevant, for instance, if we consider an experimental set-up as the realization of an individual unstable system with specific values for certain parameters and ask whether we can expect the results of the experiment to be at least qualitatively the same if we repeat it with another realization of the set-up which will presumably incorporate slightly different values of the parameters. If the system belongs to a stable family, we can expect similar results for the bifurcation behaviour.)

What I said above about the motivation for connecting robustness and explainability in the case of individual systems equally applies to the extended notion of structural stability. The existence of a stable family, even though individual members may exhibit bifurcations, is less

in need of explanation than the existence of an unstable family which, under the slightest perturbations, turns into a qualitatively different family. Scientists make frequent use of unstable families of dynamical systems (models) in order to describe systems in nature. Nevertheless, the majority opinion about this practice appears to be that even though such unstable models are useful, their unstable features (e.g., certain types of strange attractors) cannot correspond to “real” features of the systems modelled. As Guckenheimer and Holmes (1983, 259) put it, “the only properties of a dynamical system... which are physically relevant are those which are preserved under perturbations of the system.”¹⁴

In order to accommodate the existence of instabilities in nature, the characterization of grounded (or explainable) supervenience relations has to take the form of a disjunction:

A supervenience relation between property classes in a system is *grounded* (or explainable) if

- (i) the relation is structurally stable or robust; or
- (ii) if the relation is unstable, the instability occurs in a stable way, i.e., the system belongs to a family of systems which, as a family, is structurally stable.

If the wayward atom world is possible, then we have a “rather bizarre kind of dependence”, and the strangeness of the case consists in the violation of the robustness requirement for supervenience relations. Distinguishing clause (i) from clause (ii) in the requirement, we can say that the physicalist can accept strange dependencies if they violate

only (i); if they violate (ii) as well, the supervenience relations are not ‘explainable’ and should be taken as a refutation of physicalism.

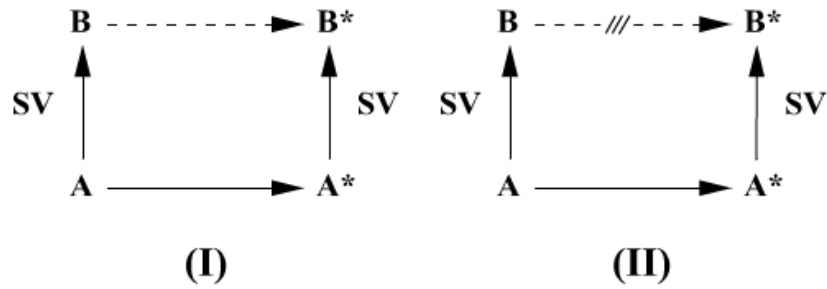
One could perhaps try to show — although I cannot argue for it here — that a violation of clause (ii), an ‘unstable instability’, amounts to a violation of a necessary feature of laws of nature, thus rendering a case of unstable instability nomologically impossible. That is, if we suppose the laws of nature are something like the structurally stable features (in sense (ii)) of the world¹⁵, it could be argued that the wayward atom world is incompatible with the laws *if* the scenario indeed violates condition (ii). Any version of the supervenience notion, like, e.g., Hellman and Thompson’s (1975), that uses nomological instead of wider notions of possibility, would then automatically exclude the wayward atom case.¹⁶

4. Emergence

Whenever a supervenience relation violates the requirement of structural stability but satisfies the second condition of groundedness — that the instability occurs in a stable way —, I suggest we have a potential case of *diachronic* property emergence: a slight change in the base properties $A \rightarrow A^*$ can lead to supervening properties B^* which are qualitatively different from the properties B that supervened on the unmodified base A . Emergence, in the diachronic sense, is not understood in this framework as a violation or breakdown of the supervenience relation itself. Supervenience relations are (synchronic) relations between base

and supervening properties existing at the same time. The emergence relation to be defined here, by contrast, is a (diachronic) relation between systems at successive times.¹⁷ The notions of supervenience and emergence become connected, however, through the requirement of robustness for the supervenience relation. Robust supervenience adds a counterfactual condition to supervenience: it specifies what would happen were we to subject the system to a temporal evolution of the base properties, i.e., a certain change in the base would result in a certain modification of the supervenient properties which we then compare to the unmodified supervenient properties. If the modification is radical enough, the robustness requirement is violated (but supervenience may still hold). For diachronic emergence we need an actual, not counterfactual, change in the base resulting in a qualitative change of the supervening properties. A violation of the first clause of the robustness requirement for the supervenience relation is therefore a necessary but not a sufficient condition for diachronic emergence. Violation of robustness says what would happen if there were a modification of the base; diachronic emergence describes what in fact happens in that case.

Let $B \rightsquigarrow B^*$ refer to the temporal evolution of a system's behaviour at one time (B) into the behaviour at a later time (B*) such that topological equivalence is respected, and let $B \rightsquigarrow\!\!\!\!/\!\!\!\! \rightsquigarrow B^*$ refer to evolution into inequivalent properties. We then have the schemas shown in Figure 5 (SV is the supervenience relation):



(FIGURE 5)

If we read the the evolution of base and supervening properties *counterfactually*, (I) represents robust supervenience and (II) non-robust supervenience of B on A. Reading the diagrams as indicating *actual* evolution, (I) describes the case of (diachronically) non-emergent properties B* and (II) represents the case of (diachronically) emergent B*.

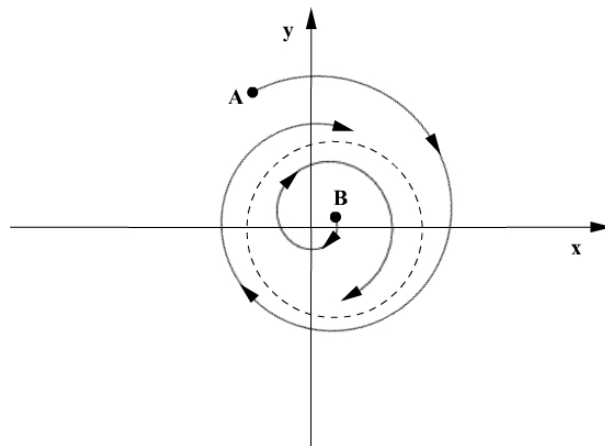
As an illustration, take as our system a damped *nonlinear* oscillator, realized, for instance, by a triode circuit involving a nonlinear resistor. Let the amplitude of the oscillations, $x(t)$, be given by van der Pol's equation

$$\begin{aligned} dx/dt &= y \\ dy/dt - \eta(1 - x^2)y + x &= 0 \end{aligned}$$

where η is a parameter measuring the strength of the damping factor $(1-x^2)y$. For large amplitudes x this factor is negative, thus keeping the oscillations bounded; for small

amplitudes the factor becomes positive and thereby excites the oscillatory movement.¹⁸

Suppose the damping could be reduced to zero ($\eta = 0$). We would then have an undamped harmonic oscillator with the familiar phase portrait of a system of concentric ellipses (Figure 3). From whatever initial conditions (x, y) you start the system, it will oscillate such that it periodically passes through these same conditions. In the language of dynamical systems theory: the system has as its equilibrium point a 'center' at the origin ($x = 0, y = 0$). If we gradually turn on the damping, the system will still show oscillations (Figure 6):



(FIGURE 6)

Now, however, the oscillations have a 'limit cycle', that is, from wherever you start the oscillator (arbitrary initial conditions A and B), the system will tend towards a unique periodic behaviour, the cycle shown in the figure as a dashed closed curve.

Clearly, this system has the same basic determinables in the undamped and in the

damped regime, in particular, position (x) and momentum (y). Only the succession of values which these determinables take on changes. (The fact that the damping increases from zero to a finite value must not be taken as indication that a new determinable (viz., damping) has been introduced into the system. Nothing hangs on the particular value of zero for η . The kind of qualitative change in behaviour we are interested in occurs in general at a ‘critical value’ of a parameter, whether this is zero or some other value.) Although no new determinables have been introduced into the supervenience base of the system compared to the base with $\eta = 0$, the limit cycle behaviour of the oscillator in the damped regime is a *novel* property, a feature which distinguishes the system with damping *qualitatively* from the system without damping.

This feature of novelty of a property is a traditional ingredient in the notion of emergent properties and is captured in our framework by the fact that the phase space portraits in Figures 3 and 4 are topologically inequivalent: no smooth deformation will transform the set of trajectories of Figure 3 into those of Figure 4, even though, for small enough damping, the trajectories of the damped system can be *quantitatively* very close to those of the undamped system. In other words, the oscillator is structurally unstable around the critical value of the control parameter η ; small variations of the parameter (perturbations of the system) will lead to qualitatively different behaviour. At the critical value of η the system undergoes a bifurcation.

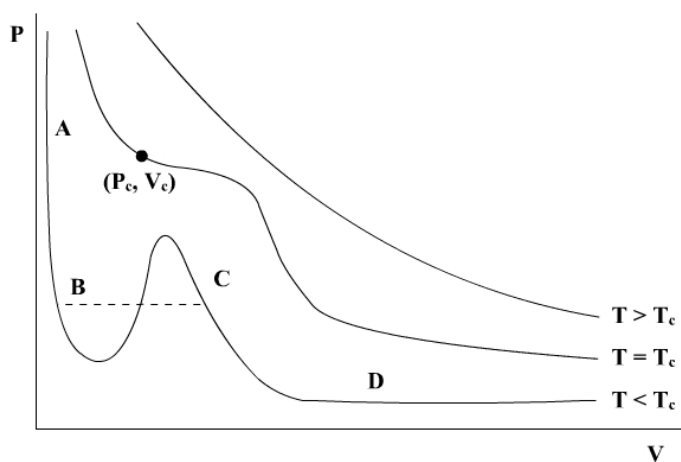
While the first disjunct of the definition of robust supervenience is thus clearly violated

for a realization of the van der Pol oscillator with a value of η near the bifurcation value, we still have to ensure that the second disjunct condition is satisfied as well. Indeed, if we consider a *family* of equations describing van der Pol oscillators, the instability (the formation of a limit cycle) does occur in a stable way; i.e., it does not disappear under small perturbations of the family (cf., e.g., Seydel 1994, chs.2.10 and 8). We therefore have a case of a novel property emerging in a system without a breakdown of supervenience.

A further illustration of the notion of emergence suggested here is provided by the phenomenological description of *phase transitions* in non-ideal gases. A rough approximation to the equation of state of such gases is van der Waals' equation, connecting pressure P , volume V , and temperature T :¹⁹

$$(P + a/V^2) (V - b) = RT$$

This is formally a cubic equation in V . Therefore, there will be three qualitatively different kinds of isotherms (pressure as a function of volume at constant temperature), depending on the number of real roots possessed by the equation for V (Figure 7). For large T , the isotherms are of qualitatively the same character as those for the ideal gas: hyperbolas, with V monotonically decreasing with increasing P . For small T , the isotherms have a maximum as well as a minimum, thus they are no longer monotonically decreasing. Separating these two cases we have a critical isotherm with $T = T_c$ which has a point of inflection at (P_c, V_c) .



(FIGURE 7)

A phase transition from liquid to gas can be described in this model qualitatively as follows: A *liquid*, under decreasing pressure (along the line A - B) will not change its volume much. At a certain value of P , however, the liquid turns into a gas; the system ‘jumps’, at constant pressure, from volume V_2 to V_1 (along the horizontal line B - C),²⁰ further decrease in pressure now leads to a large increase in volume (along C - D), typical of *gases*. This jump to a phase with qualitatively new properties can happen only if the isotherm has a minimum as well as a maximum.

If T is taken as our parameter, we can see a bifurcation occurring at $T = T_c$ when the ‘shape’ of the isotherm changes from having no extrema to having minima and maxima. The inflection point (P_c, V_c) splits into two extrema. This splitting corresponds to generating the possibility of phase transitions. Small variations of T far away from the critical value result only in small quantitative — as opposed to qualitative — changes in the P - V characteristics. Around T_c , however, small changes in T can lead to qualitatively different isotherms, that is, around T_c , we can find isotherms which are arbitrarily close in their T -values but qualitatively different in the phase properties they describe. (The two functions, $p = p_{T_c}(V)$ and $p = p_{T < T_c}(V)$, are not topologically equivalent, i.e., they cannot be related by a homeomorphism in the way described above. In other words: the change introduced by varying the parameter T around T_c cannot be compensated by a coordinate transformation.) As in the case of the van der Pol oscillator, the bifurcation in the van der Waals’ equation occurs in a stable way, i.e., it does not get removed under small perturbations of the underlying family of equations.²¹

We then have the following picture for the emergence of a novel property: A system with base properties A , characterized by a value of a control parameter p , under slight variation of the parameter around its bifurcation value, turns into a system with base properties A^* , which shows qualitatively different behaviour B^* than the original system with behavioural characteristic B . The original system with $\{A, B\}$ I call the *reference system* for the system with $\{A^*, B^*\}$. A property $\mathbf{b} \in B^*$ in a system $\{A^*, B^*\}$ is emergent if

- (1) reference systems $\{A, B\}$ with bases A different from A^* only in the value of the control parameter p , do not have \mathbf{b} ; and
- (2) the behaviour of $\{A^*, B^*\}$ in which \mathbf{b} manifests itself (in the phase space portrait) is qualitatively different from, or topologically inequivalent to, the behaviour of $\{A, B\}$.

Thus, the novel property of the van der Pol oscillator with $\eta > 0$ considered above would be ‘having a limit cycle’, a feature missing from the oscillator with $\eta = 0$. In the case of the non-ideal gas (van der Waals’ equation), the novel property of the system below T_c would be the existence of a phase transition, which is missing in systems above the critical temperature.

Note, again, that this relation of (diachronic) emergence is *not* defined, in contrast to the supervenience relation, between the B-level properties and the A-level properties (at the same value of the parameter) but rather between different systems $\{A, B\}$ and $\{A^*, B^*\}$.²² Another important feature to note is that the relation of being qualitatively different, or topologically inequivalent, is symmetric. If the pre- and post-bifurcation behaviour of a system are qualitatively different from each other, we have a symmetric relation: each behaviour could equally well be called ‘novel’ with respect to the other. Why then should we regard the radical modification of a system’s behaviour upon variation of a control parameter ‘novel’ or ‘emergent’?

This problem is similar to a well-known problem about the supervenience relation: supervenience is not, as it intuitively should be, an asymmetrical relation but merely non-symmetric: B supervenes on A is compatible with A supervenes on B. From this perspective, for instance, physicalism is based on a prejudice in favour of the physical, not on a feature of the supervenience relation itself. But while this symmetry problem would clearly affect an account of *synchronic* emergence in terms of topological inequivalence, it obviously does not apply to *diachronic* emergence as defined here: since the properties we compare are not simultaneous but temporally successive properties, the direction of time automatically introduces the desired asymmetry.

5. Non-Reducibility

The appearance of novelty which is implied by our notion of emergence can also be shown to be connected with a sense of non-reducibility of the new property to the properties of the reference systems. Non-reducibility (in different versions), besides novelty, has been another traditional ingredient in the notion of emergent properties and our framework will allow us to reconstruct this feature.²³

Consider van der Waals' equation again. Is the system's property of having a phase transition (liquid \rightarrow gas) at $T < T_c$ *reducible* to the properties that the system possesses at $T > T_c$? The question of reducibility will be phrased in terms of descriptions of, or theories

about, the properties of the systems and a notion of reduction as a relation between such descriptions or theories has to be chosen. We take a fairly liberal approach, not as demanding as deducibility of one theory from another (as in Nagel-type reductions). We say that a theory Θ reduces to another theory Θ' if $\lim \Theta = \Theta'$ under an appropriate choice of parameter p for the limit. Thus, the solutions to the equations of motions of Θ , in the limit of $p \rightarrow 0$, should coincide with the solutions of the equations of Θ' where we set $p = 0$. Special Relativity Theory (Θ), for example, reduces to Classical Mechanics (Θ') because Θ goes smoothly over into Θ' in the limit of $v/c \rightarrow 0$.²⁴ Note that if Θ is not reducible to Θ' in this sense, then there will also be a failure of reduction in the sense of Θ' being deducible from Θ . The question in our case then is: Can a description of the van der Waals system below the critical temperature (corresponding to Θ) be reduced to a description of the system above T_c (corresponding to Θ'), i.e., does Θ smoothly go over into Θ' in the limit $T \rightarrow T_c$?

The answer for the van der Waals system is simple. Take the volume of the gas as the variable that depends on varying the pressure and parameterize the family of curves with T . For $T > T_c$, the dependence of V on P is single-valued, that is, $V(P)$ is a function which assigns a unique V value to each value of P . At the critical temperature, the relation V - P is still continuous but no longer differentiable: $(\partial V/\partial p)_{T=T_c}$, which measures the compressibility of the system, diverges. For $T < T_c$, finally, V is no longer a function of P since several different V values can now be assigned to the same value of P . As we approach the critical

temperature (associated with the critical pressure P_c), V can no longer be approximated by a power series in $(P - P_c)$: small changes in P around P_c are no longer connected with ‘commensurate’ changes in V . Our reducibility requirement $\lim \Theta = \Theta'$, for some appropriate parameter, means that Θ' , in the limit of the parameter, approximates Θ arbitrarily closely. Obviously, for our case, this condition is not satisfied: The single-valued function $V(P)$ cannot continuously approximate the multi-valued relation V - P . The two descriptions of the system for different parameter regimes are thus not connected through a continuous limit relation; we cannot understand Θ , around the critical value of the parameter, as Θ' plus small corrections as we do in the case of Special Relativity and Classical Mechanics. In other words, the description for $T < T_c$ does not reduce to the description for $T > T_c$.

Such discontinuous limit relations between theories are much more common than one might have expected from the fact that this type of relation has been largely ignored by philosophers of reduction.²⁵ Harold Grad, one of the pioneers of mathematically rigorous studies of the thermodynamic limit, pointed out long ago that the relations between descriptions of the same phenomena at different levels (e.g., micro- and macro-level theories of gases) can “always be expected to be subtle and mathematically singular [i.e., discontinuous].” (Grad 1967, 54) In such cases, even though there is one overall theory which covers the two regimes, we have to choose different descriptions or descriptive levels for different regions of the parameter space without the descriptions being reducible to each

other. When the transition from one parameter regime to another is a temporal evolution, we have a case of diachronic emergence in our sense — a sense which captures two intuitions traditionally underlying the concept, *viz.*, novelty and irreducibility of emergent properties.²⁶

REFERENCES

Arnold, Vladimir I. (1983), *Geometrical Methods in the Theory of Ordinary Differential Equations*. New York: Springer.

Bailey, Andrew (1999), “Supervenience and Physicalism”, *Synthese* 117: 53-73.

Batterman, Robert W. (1995), “Theories Between Theories”, *Synthese* 103: 171-201.

Bechtel, William and Robert C. Richardson (1992), “Emergent Phenomena and Complex Systems”, in Ansgar Beckermann et al. (eds.), *Emergence or Reduction?* Berlin: De Gruyter, 257-288.

Berry, Michael (1994), "Asymptotics, Singularities and the Reduction of Theories", in Dag Prawitz et al. (eds.), *Logic, Methodology, and Philosophy of Science IX*. New York: Elsevier, 597-607.

Chen, Lin-Yuan, Nigel Goldenfeld, Y. Oono, and Glenn Paquette (1994), "Selection, Stability and Renormalization", *Physica A* 204, 111-133

Grad, Harold (1967), "Levels of Description in Statistical Mechanics and Thermodynamics", in Mario Bunge (ed.), *Delaware Seminar in the Foundations of Physics*. Heidelberg: Springer, 48-76.

Guckenheimer, John and Philip Holmes (1983), *Nonlinear Oscillators, Dynamical Systems, and Bifurcations of Vector Fields*. New York: Springer.

Hellmann, Geoffrey and Frank Thompson (1975), "Physicalism: Ontology, Determination, and Reduction", *Journal of Philosophy* 72, 551-564.

Horgan, Terence (1993), "From Supervenience to Superdupervenience", *Mind* 102, 555-586.

Humphreys, Paul (1996), "Emergence, Not Supervenience", *Philosophy of Science* 64, Supplement (*PSA 1996, Vol.II*), S337-S345.

Hüttemann, Andreas and O. Terzidis (forthcoming), "Emergence in Physics", *International Studies in the Philosophy of Science*

Kim, Jaegwon (1984), "Concepts of Supervenience", in Kim 1993, 53-78.

___ (1987), "'Strong' and 'Global' Supervenience Revisited", in Kim 1993, 79-91.

___ (1989), "The Myth of Nonreductive Materialism", in Kim 1993, 265-284.

___ (1990), "Supervenience as a Philosophical Concept", in Kim 1993, 131-160.

___ (1992), "'Downward Causation' in Emergentism and Nonreductive Physicalism", in Ansgar Beckermann et al. (eds.), *Emergence or Reduction?* Berlin: DeGruyter, 119-138.

___ (1993), *Supervenience and Mind*. New York: Cambridge University Press.

____ (1997), "The Mind-Body-Problem: Taking Stock After Forty Years", *Philosophical Perspectives* 11, 185-207.

McLaughlin, Brian (1995), "Varieties of Supervenience", in: Elias E. Savellos et al. (eds.), *Supervenience. New Essays*. New York: Cambridge University Press, 16-59.

Mormann, Thomas (1994), "Accessibility, Kinds, and Laws: A Structural Explication", *Philosophy of Science* 61, 389-406.

Newman, David V. (1996), "Emergence and Strange Attractors", *Philosophy of Science* 63, 245-261.

Nickles, Thomas (1973), "Two Concepts of Intertheoretic Reduction", *Journal of Philosophy* 70, 181-201.

Paull, R.C./Theodor R. Sider (1992), "In Defense of Global Supervenience", *Philosophy and Phenomenological Research* 52, 833-854.

Port, Robert F. and Timothy Van Gelder (eds.) (1995), *Mind as Motion*. Cambridge, Mass.: MIT Press.

Poston, Tim and Ian Stewart (1978), *Catastrophe Theory and its Applications*. London: Pitman.

Rueger, Alexander (forthcoming), "Physical Emergence, Diachronic and Synchronic", *Synthese*

Rueger, Alexander and W.David Sharp (1998), "Metaphysical Presuppositions of Scientific Practice: Atomism Vs. Wholism", *Canadian Journal of Philosophy* 28, 1-20.

Saunders, Peter T. (1980), *An Introduction to Catastrophe Theory*. Cambridge, England: Cambridge University Press.

Seydel, Richard (1994), *Practical Bifurcation and Stability Analysis*. New York: Springer.

Wilson, Mark (1985), "What Is This Thing Called 'Pain'?" *Pacific Philosophical Quarterly* 66, 227-267.

Wimsatt, William (1996), “Aggregativity: Reductive Heuristics for Finding Emergence”, *Philosophy of Science* 64, Supplement (*PSA 1996, Vol.II*), S372-S384.

FOOTNOTES

-
1. For accounts of these developments, see Horgan 1993 and McLaughlin 1995.
 2. See Hüttemann and Terzidis (forthcoming) for more detailed criticism along these lines.
 3. It is this second clause which, apart from the technical framework of dynamical systems theory employed here, distinguishes robust supervenience from Kim’s similarity-based supervenience.
 4. Cf., however, the discussion in Wilson 1985 about whether any mental properties could possibly fall outside the scope of dynamical systems theory. The ‘dynamical theory of the mind’ (Port and van Gelder 1995) is an attempt to extend this framework into the mental domain.
 5. For an informal introduction to structural stability see, e.g., Saunders 1980, 17-21.
 6. This is to be distinguished from variations of the initial conditions of the system which leave the dynamics itself unchanged.
 7. Adapted from Arnold 1983, ch.3. Cf. also Guckenheimer and Holmes 1983, 38f.

-
8. Usually one postulates that a map f is *close* to a map g if g belongs to an ε -neighbourhood of f such that every map in that neighbourhood agrees with f and its derivatives up to $\varepsilon > 0$.
 9. Supposing we allow the scenario as possible, i.e., we grant the existence of a supervenience relation in this case.
 10. That's why Newman's (1996) attempt to develop a theory of emergence on the basis of a system's behaviour being sensitively dependent on initial conditions does not seem to capture the sense in which emergent properties are supposed to be 'novel'.
 11. Cf. Rueger and Sharp 1998 for further illustrations.
 12. For unfoldings and their classification into the 'elementary catastrophes', see, e.g., Poston and Stewart 1978, chs. 7-9. Chen et al. (1994) have argued for a restriction of 'all possible small perturbations' to 'all physically small perturbations'. (Thanks to Bob Batterman for this reference.)
 13. Adapted from Arnold 1983, ch.6.
 14. See Rueger and Sharp 1998 for further discussion (structural stability of families of systems is there labelled 'structural stability₂'). See also the interesting applications of this view in the work of Goldenfeld, Oono et al. (e.g., Chen et al. 1994).
 15. Cf. Mormann (1994) for a suggestion along such lines.

16. One could still doubt that the notion of robust supervenience suggested here is of much help in the task of characterizing physicalism. It seems easy to imagine a scenario which satisfies the robustness criterion but nevertheless is unacceptable to a physicalist, thus indicating that, for physicalists, robustness is the wrong feature to require of a supervenience relation. Suppose the distribution of supervenient mental properties M in the world is so stable against changes in the physical base P that the M properties exist regardless of whether P is there or not. Suppose, that is, that evaporating my brain, a perturbation of the base properties, leaves my mental states, the M level, qualitatively unchanged. This scenario is compatible with a bare supervenience relation of P and M; furthermore, the relation is, presumably, structurally stable. Doesn't a case like this show that adding robustness to the supervenience relation, far from assisting in the characterization of physicalism, is incapable of ruling out all kinds of possibilities — like dreaming corpses and theistic occasionalism — that no physicalist could be happy with? (For a recent discussion of this and related problems cf. Bailey 1999.)

Remember, however, that the definition of structural stability requires that the system in question (i.e., the system relating P and M at some time) be topologically equivalent to *nearby* systems, that is, systems which are related to the original system through a *small* perturbation. The perturbation which leads from the system where I have my brain intact to the system where it has evaporated, however, is clearly to be regarded as a major, rather than a small, change in the original system. What may or may not happen under such large perturbations is not covered by

the notion of structural stability; that the system retains its M features under large changes in P neither shows nor disproves, that the system is robust. Evaporating my brain while my mental properties stay in place would not by itself imply (or rule out) that the relation of my P and M properties, although compatible with bare supervenience, is structurally stable.

17. For the case of *synchronic* emergence, see Rueger (forthcoming).
18. For a discussion of the van der Pol equation, cf. Guckenheimer/Holmes 1983, ch.2.
19. For a discussion of van der Waals' equation cf., e.g., Poston and Stewart 1978, 327ff.
20. On most of the curved connection between B and C, $(\partial P/\partial V)_T > 0$ which means that this path is unstable since small increases in volume would lead to increases in pressure which would in turn lead to further increase in volume, etc.

21. The van der Waals equation can be transformed into the standard equation for the 'cusp catastrophe' which is structurally stable. See Poston and Stewart 1978, 329.

22. An alternative but closely related way of phrasing requirements for emergence is Wimsatt's criterion of "violation of aggregativity" (Wimsatt 1996; cf. also Bechtel and Richardson 1992). For details see Rueger (forthcoming).

23. More on non-reducibility in Rueger (forthcoming).
24. For this notion of reduction and its relation to Nagel-type reductions, cf. Nickles 1973.

-
25. The exception to this is the pioneering work of Batterman, e.g., 1995. Cf. also Berry 1994.
 26. The notion also seems to capture most of the desiderata Humphreys (1996) spells out.