Supervenience and Emergence

The metaphysical relation of supervenience has seen most of its service in the fields of the philosophy of mind and ethics. Although not repaying all of the hopes some initially invested in it – the mind-body problem remains stubbornly unsolved, ethics not satisfactorily naturalized – the use of the notion of supervenience has certainly clarified the nature and the commitments of so-called non-reductive materialism, especially with regard to the questions of whether explanations of supervenience relations are required and whether such explanations must amount to a kind of reduction.

I think it is possible to enlist the notion of supervenience for a more purely metaphysical task which extends beyond the boundaries of ethics and philosophy of mind. This task is the clarification of the notions of emergence and emergentism, which latter doctrine is receiving again some close philosophical attention (see McLaughlin, Kim ??).

I want to try to do this in a 'semi-formal' way which makes as clear as possible the relationships amongst various notions of supervenience as well as the relationship between supervenience and emergence. And I especially want to consider the impact of an explicit consideration of the temporal evolution of states – an entirely familiar notion and one crucial to science and our scientific understanding of the world – on our ideas of supervenience and, eventually, emergence for these are significant and extensive. I do not pretend that what follows is fully rigorous, but I do hope the semi-formality makes its commitments and assumptions clear, and highlights the points of interest, some of which I think are quite surprising.

§1. Definitions.

A theory, T, is *total* if and only if it possesses completeness, closure and resolution. These are jointly defined as follows: Completeness is the doctrine that everything in the world is a T-entity or, in principle, has a non-trivial T-description and as such abides by closure and resolution.

Closure entails that there are no 'outside forces' – everything that happens, happens in accordance with fundamental T-laws so as to comply with resolution. Resolution requires that every process or object be resolvable into elementary constituents which are, by completeness, T-entities and whose abidance with T-laws governing these constituents leads to closure. For the particular example of physics (the only theory that could have any chance of being total) these definitions become: Completeness is the doctrine that everything in the world is *physical* (has a non-trivial physical description[1]) and as such abides by closure and resolution. Closure entails that there are no 'outside forces' – everything that happens, happens in accordance with fundamental *physical* laws so as to comply with resolution. Resolution requires that every process or object be resolvable into elementary constituents which are, by completeness, *physical* and whose abidance with physical laws governing these elementary constituents leads to closure[2].

We could distinguish a merely 'formal' notion of totality from that defined above. A 'formally total' theory is one that would be total if only it were true. Thus it is arguable that Newtonian mechanics is formally total (but had no chance of being true of this world). Since we are going to *assume* that final-physics – whatever it may turn out to be – is true, the notions of formal totality and totality collapse for it.

A world, W, is total if and only if a true theory of W is total.

T-possibility: something is T-possible if and only if it exists in some T-possible world, that is, some world that obeys the fundamental laws of T. (Example: physical possibility is existence in some physically possible world, that is, a world that obeys the fundamental laws of physics. To avoid making physical possibility epistemically relative we can regard physics to be the true, *final* physics – whether or not humans ever manage to discover such a theory. We can call this 'final-physical-possibility'.)

[1] 'Non-trivial' is added here and above to prevent properties like 'having charge +1 or not' rendering anything and everything a physical entity ($\sqrt{2}$ has this property).

[2] It may be worth noting here that this is not an endorsement of 'part-whole reductionism', though it is consistent with it. We know from quantum mechanics that the states of 'wholes' are not 'simple' functions of the states of their parts but this does not tell against the characterization given in the text. Quantum mechanics is a *celebration* of how the fundamental interactions of things can be understood – rigorously understood – to yield new features. It is, if you like, a mathematically precise theory of emergence, but one that obeys the strictures of resolution. The kind of emergence endorsed by quantum mechanics is what will be called below *benign* emergence.

Efficacy: a state, F, of system σ is *efficacious* in producing a state, G, in system π if and only if had σ not been in state F, π would not have been in state G (it is possible, and usually the case, that σ = π). F has *efficacy* if and only if there is a state for which F is efficacious in its production.

T-Efficacy: a state, F, of system σ is *T-efficacious* in producing a T-state, G, in system π if and only if had σ not been in state F, π would not have been in state G (it is possible, and usually the case, that σ = π; the state F may or may not be a T-state). F has *T-efficacy* if and only if there is a state for which F is T-efficacious in its production.

Supervenience. Supervenience is a special relation of dependence of one domain upon another. It is often taken to be a relation between families of properties or states, where a 'family' of properties is a set of properties that define the domains at issue (for example, the psychological properties would form one family while the physical properties would form another, and we would claim that there was a supervenience relation of the psychological properties upon the physical properties if, in accordance with the special relation of dependence defining supervenience, we held that (all instances of) psychological properties depended upon (instances of) physical properties). It is natural to extend the notion to speak of the supervenience of one theoretical domain upon another, in which case the state or property families are given by the theories at issue (as it might be, psychology versus physics). We could also allow a supervenience relation between theories by extension of a supervenience relation between the theoretical domains (even where these domains might be hypothetical rather than actual).

The nature of the dependence relation which defines supervenience is intentionally rather unspecified, but one core idea is that there can be no difference in the supervening domain without a difference in the subvening domain (for example, we might claim that there can be no psychological difference without an underlying physical difference). A natural way to express this is in terms of indiscernibility with respect to the subvening domain requiring indiscernibility with respect to the supervening domain. Another way is to define supervenience directly in terms of the determination of properties in the supervening family by the properties of the subvening family. It is interesting that these two approaches lead to very distinct forms of supervenience.

Let's begin with a catalogue of some basic forms of supervenience, after which I want to

introduce a new form that comes in several variants. The three basic forms of supervenience of interest here are strong, weak and global supervenience (see Kim, Haugeland, Seager ??). The former two notions of supervenience can be formally expressed in terms of families of properties and a direct relation of determination between them. In what follows I will take the line that property families are given by the distinctive features employed by particular theories. The family of chemical properties is the set of properties distinctively used by chemistry, the family of physical properties is that set of properties distinctively utilized by physics, etc.. I add the term 'distinctively' only to indicate that there must be some selection from all properties mentioned by a theory since some are inessential to that theory. We can also expect that there might be some overlap between theories, but I think we ought to regard this common occurrence as an intrusion of one theoretical scheme into another. In such cases, we ought to assign the overlapping property to the more basic theory.

Given a pair of families of properties, we can define a supervenience relation between them in various ways. Strong supervenience is defined as

(SSUP) Strong Supervenience: Property (or state) family U strongly supervenes upon family T if and only if $\Box(\sigma)(F \in U)(Fx \supset (\exists G \in T)(G\sigma \,\&\, \Box(\pi)(G\pi \supset F\pi)))$

This says that it is necessarily true that for any instance of a property in B there is a property in A such that having that property guarantees having the B property. It does not say, though it permits, that some particular A property underlies every instance of the target B property. Instead, it is typically thought that there can be what is called 'multiple realization', in which a variety of A properties subvene the instances of the target B property. Notice the second necessity operator, which ensures that G subvenes F in every possible world. (That is, in every possible world, anything that manages to exemplify G will also exemplify F, but not necessarily vice versa.)

One should wonder about the nature of the necessity deployed in this definition, as well as the ones to follow. If one claims that U supervenes upon T, then what is at issue is whether in every T-possible world, we have the appropriate relation between the properties of U and T. What happens in T-impossible worlds is irrelevant. More concretely, physicalists expect that high-level features of the world, as dealt with by theories such as chemistry, biology, psychology, etc. will supervene upon physical features. They need not concern themselves with what happens in

possible worlds that are *physically* impossible, such as worlds where only ectoplasmic spirits exist and where supervenience upon the physical could hardly be expected to hold. Of course, there is then the issue of whether we live in a world that is physically possible (i.e. a world that obeys the fundamental laws of physics). Physicalists think we do live in such a world, and it is hard to doubt that there is a good deal of evidence in support of this position.

The second form of supervenience of interest to us is defined as follows:

(WSUP) Weak Supervenience: Property (or state) family U weakly supervenes upon family T if and only if $\Box(\sigma)(F \in U)(F\sigma \supset (\exists G \in T)(G\sigma \,\&\, (\pi)(G\pi \supset F\pi)))$

The difference between weak and strong supervenience, intuitively speaking, is that although they agree that the supervening domain is determined by states of the subvening domain, this determination can be different in different worlds. A simple (if somewhat imperfect) example, originally due to Jaegwon Kim I believe, of weak supervenience is the supervenience of the truth of a sentence upon the sentence's syntax. It must be that any two sentences that are syntactically identical have the same truth value (and of course every sentence has a syntactic structure)[3]. But we do not expect the truth value to be the same from world to world, as we vary the facts which make the sentences true. We might thus expect that syntactic structure plus a specification of the facts *strongly* subvenes the truth of sentences. The difference between weak and strong supervenience will turn out to be very important for the clarification of various notions of *emergence*.

A quite difference approach to supervenience is also possible. One can express supervenience in terms of indiscernibility rather than property determination. One method of doing this is in terms of possible worlds and thus avoids the explicit appeal to modal operators. Supervenience of U upon T would require, at least, that in every T-possible world, if there is agreement about the assignment of T-states to systems then there is agreement about the assignment of U-states to systems. We might write this as:

(GSUP) Global Supervenience: $(w)(w^*)(\sigma)(F \in U)((w =_T w^* \,\&\, F\sigma w) \supset F\sigma w^*)$,

where w and w* are T-possible worlds, σ ranges over systems and F over U-states. The modified

[3] An obvious imperfection I gloss over is the existence of indexical terms. With a little suppression of one's critical, or is it pedantical, faculties the point of the example should be clear.

identity symbol, '$=_T$' is short for 'is identical in T respects to'. In order to allow for a non-trivial role for the temporal evolution of states I am going to modify this standard definition of global supervenience by requiring that the T-indiscernibility of worlds be restricted to indiscernibility up to the time when $F\sigma w$ obtains. This entails that global supervenience, as defined here, can fail for properties that depend for their existence at a time on states which occur later *if* worlds lack what I will call below T-temporal supervenience (or temporal determination). Some properties do have this dependence upon the future. Whether a prediction – a future tensed sentence – is *true* or not obviously depends upon the future. Less trivially, whether an action, for example, is good or bad might depend upon its consequences. If two worlds which were T-indiscernible up to the time of the action could diverge (with respect to T) after the action then it could be that the action was good in one world but bad in the other[4]. If we were inclined to hypothesize that moral properties and 'consequences' supervene upon the physical state of the world up to the time of the action such divergence would represent the failure of that hypothesis. We could distinguish an 'absolute' global from a 'limited' global supervenience of U upon T, the former involving absolute world T-indiscernibility across all space and time, the latter only indiscernibility up to the occurrence of a given U-state. Fortunately, such a distinction would be of little assistance in what follows, so I shall resist adding yet another kind of supervenience.

In any event, this formulation reveals an ambiguity in the notion of supervenience (see Kim, Haugeland, Petrie, Seager ??). The formulation of global supervenience in terms of worlds, unlike the definition of strong supervenience given above, does not explicitly require that the T-state that subvenes a U-state be a state of the very same system that exemplifies the U-state. This is thus a very weak form of supervenience. For example, it permits two worlds that differ only in the position of a single atom somewhere in, say, the star Vega to have radically distinct distributions of psychological properties – perhaps one world is ours but in the other there are no minds whatsoever!

It is not difficult to strengthen GSUP to a form that makes the indiscernibility of particular systems rather than whole worlds the basis of supervenience, a form we might call local

[4] Actually, this raises the interesting issue of whether in principle unpredictable consequences should be considered as moral pluses or minuses in the assessment of actions as they accrue, as opposed to evaluating actions in terms of *present* expected utility, or the expected utility of the action at the time it was performed.

supervenience:

(LSUP) Local Supervenience: $(w)(w^*)(F \in U)(\sigma)(\pi)(((G \in T)(G\sigma w \equiv G\pi w^*) \& F\sigma w) \supset F\pi w^*)$[5].

This adds the condition that it is the systems, σ and π, that are such that if they are T-indiscernible across possible worlds then they will also be U-indiscernible. Local supervenience is not quite the same as strong supervenience[6]. The latter does not require full local indiscernibility as a condition of supervenience but only the sharing of one critical property from the subvening family. Though less weak than GSUP, LSUP is still a very weak notion of supervenience. It permits, for example, two systems which differ only in the position of an atom in their big toe to differ radically in their psychological properties – perhaps one system is me but the other has no psychological properties at all!

Problematic examples such as that of Vega or the big toe reinforce the intuitively plausibility of the 'super-localization' of strong supervenience, for it seems reasonable to suppose that some T-properties might be irrelevant to the possession of U-properties. For example, in some possible worlds (maybe even in the actual world) there are creatures physically identical to us except that they are made out of anti-matter rather than matter. This would seem to be psychologically irrelevant but they would fail the test of indiscernibility since although systems composed of matter and anti-matter can share almost all their physical properties they are obviously physically discernible. Of course, global and local supervenience do not *prevent* non-identical systems from possessing the same supervening properties, but we could not use either global or local supervenience to argue for our anti-matter cousins's possession of mind, whereas strong supervenience would probably – depending upon the range of physical properties we take to subvene mind – provide such an argument.

Evidently, strong supervenience implies local supervenience but not vice versa. If we assume strong supervenience and the antecedent of local supervenience we obtain the local T-indiscernibility of σ and π across w and w*. By strong supervenience, there is a T-state, G, that σ

[5] A 'weak' local supervenience is expressed in terms of worlds as: $(w)(\sigma)(\pi)(F \in U)(((G \in T)(G\sigma w$ if and only if $G\pi w) \& F\sigma w) \supset F\pi w)$. It is a trivial consequence of LSUP.

[6] The 'direct' translation of strong supervenience into possible world terms would be the not very interesting: $(w)(\sigma)(F \in U)(F\sigma w \supset (\exists G \in T)(G\sigma w \& (w^*)(\pi)(G\pi w^* \supset F\pi w^*)))$.

has which necessitates F. Since σ and π are T-indiscernible, π must also possess G and therefore we must have $F\pi w^*$. The reverse fails for we cannot deduce simply from the fact that σ and π share G across possible worlds that σ and π are fully T-indiscernible across the worlds (unless we allow, as I think we should not, some very dubious metaphysical chicanery which encodes every feature of a possible world as a property of an individual in that world, such as 'properties' like 'exists in a world where the speed of light is 300,000 km/s').

It is, furthermore, obvious that local supervenience implies global supervenience but that once again the reverse fails to hold (since the assumption of local T-indiscernibility of two systems will not lead to the T-indiscernibility of their whole possible worlds).

However, the definitions can be brought together by fiat, if we restrict attention to domains where 'reasonable' and 'plausible' supervenience relations are local and particular. This restriction is important since it is arguable that *efficacy* is both local and dependent upon particular states, and we have a strong interest in domains that are efficacious. An illustration of a non-efficacious and non-local 'domain' is that of money. Money does not supervene locally (two locally physically identical scraps of paper could differ in their intrinsic monetary value depending upon, for example, the *intentions* and *social-status* of their creators). But for that very reason, money can't cause anything as such, but only via its exemplifying certain physical features that cause certain *beliefs* (in people) or certain other physical states (for example, in vending machines).

Temporal Supervenience (or temporal determination): The new form of supervenience I want to introduce is a 'temporal' supervenience, in which the state of a system at one time is determined by the state of the system at an another time (generally speaking, an earlier time if we think of causal determination). Temporal supervenience, as I call it, is a familiar notion with an unfamiliar name. But while it is odd to employ the term thus, I use the name 'temporal supervenience' to emphasize the analogies between the evolution of the states of systems through time and the kinds of supervenience relations we have already discussed. As we shall see, the two notions have quite deep and somewhat surprising relationships as well.

(TS) Temporal Supervenience: The states of system σ 'temporally supervene' upon the states of system σ if and only if $\Box(F)(t)(F\sigma t \supset (\exists G)(\exists t_1)(G\sigma t_1 \,\&\, (\pi)(t_2)(\exists t_3)(G\pi t_2 \supset F\pi t_3)))$

Here, and below, F and G are possible states of system s. Call F the 'successor state' and G the 'predecessor state'. To avoid clutter, it is not stated in the definitions but it is assumed that t_1 is before t and t_2 is before t_3. I make no attempt to specify the *amount* of time there should be between states or to address the issue of whether time is continuous or discrete.

(FS) Full Temporal Supervenience: The states of system σ 'fully temporally supervene' upon the states of system σ if and only if $\Box(F)(t)(F\sigma t \supset (\exists G)(\exists t_1)(G\sigma t_1 \,\&\, (\pi)(t_2)(\exists t_3)(G\pi t_2 \equiv F\pi t_3)))$

The difference between TS and FS is that in FS there is unique temporal determination both backwards and forwards in time (which is not to say that we have backwards *causation*). One can, that is, as easily foretell the past as the future of the system from its current state. Though it won't figure much in the discussion below, full temporal supervenience is nonetheless important since, generally speaking, fundamental theories of physics exemplify it.

(TTS) T/U-temporal Supervenience: The T-states of system σ 'temporally supervene' upon the U-states of system σ if and only if $\Box(F \in T)(t)(F\sigma t \supset (\exists G \in U)(\exists t_1)(G\sigma t_1 \,\&\, (\pi)(t_2)(\exists t_3)(G\pi t_2 \supset F\pi t_3)))$

(TFS) Full T/U-temporal Supervenience The T-states of system σ 'fully temporally supervene' upon the U-states of system σ if and only if $\Box(F \in T)(t)(F\sigma t \supset (\exists G \in U)(\exists t_1)(G\sigma t_1 \,\&\, (\pi)(t_2)(\exists t_3)(G\pi t_2 \equiv F\pi t_3)))$

Note that T and U can be the same theory (or family of states). In the discussion below, intra rather than inter-domain temporal supervenience will figure most prominently. So instead of writing 'T/T-temporal supervenience' I'll just use 'T-temporal supervenience'. The notions of T/U-temporal supervenience are more useful than the more basic TS and FS since we normally are concerned with the relations of temporal supervenience either within theories or across theories, rather than from an abstract, non-theoretical standpoint.

The kinds of modal differences between strong and weak supervenience can be duplicated within temporal supervenience, as follows:

(STS) Strong T/U-temporal Supervenience: The T-states of system σ 'strongly temporally supervene' upon the U-states of system σ if and only if $\Box(F \in T)(t)(F\sigma t \supset (\exists G \in U)(\exists t_1)(G\sigma t_1 \,\&\, \Box(\pi)(t_2)(\exists t_3)(G\pi t_2 \supset F\pi t_3)))$

(WTS) Weak T/U-temporal Supervenience: The T-states of system σ 'weakly temporally

supervene' upon the U-states of system σ if and only if $\Box(F \in T)(t)(F\sigma t \supset (\exists G \in U)(\exists t_1)(G\sigma t_1$ & $(\pi)(t_2)(\exists t_3)(G\pi t_2 \supset F\pi t_3)))$

(SFTS) Strong Full T/U-temporal Supervenience: The T-states of system σ 'strongly fully temporally supervenes' upon the U-states of system σ if and only if $\Box(F \in T)(t)(F\sigma t \supset (\exists G \in T)(\exists t_1)(G\sigma t_1$ & $\Box(\pi)(t_2)(\exists t_3)(G\pi t_2 \equiv F\pi t_3)))$

(WFTS) Weak Full T/U-temporal Supervenience: The T-states of system σ 'weakly fully temporally supervenes' upon the U-states of system σ if and only if $\Box(F \in T)(t)(F\sigma t \supset (\exists G \in T)(\exists t_1)(G\sigma t_1$ & $(\pi)(t_2)(\exists t_3)(G\pi t_2 \equiv F\pi t_3)))$

The differences in these definitions are exact analogues of the difference between weak and strong supervenience as given above. Intuitively, the difference is that weak temporal supervenience requires that every possible world exhibit unique state determination across time (backwards *and* forwards for full temporal supervenience) but that the particular state to state transitions can differ from world to world. This difference can matter philosophically, as we will eventually see below.

One general condition on the above definitions of temporal supervenience should be noted. It is understood that the systems in question are undisturbed systems, where undisturbed is taken to mean that there are no T-influences which are acting on the system which are not part of the system. We can allow for 'approximately undisturbed' systems where the unaccounted for T-influences are insufficient to much alter the state transitions referred to by the definitions. Also, for cases of disturbed systems, we can always 'generate' an undisturbed system by letting the boundaries of the system grow to encompass the T-disturbance.

Top-Down Discipline: A family of states (or theory), U, has Top-Down Discipline (TDD, or U/T-TDD) relative to a family of states (or theory), T if and only if

(1) U supervenes upon T[7]

(2) for every U-state, β, the set of realizer T-states is such that each element can temporally evolve into a realizer of any permitted U-successor of β.

Some discussion of the possibilities the definition allows might make this notion clearer. Assume that U supervenes upon T. TDD fails if there is a U-state, β_1, which has a set of realizers that

[7] For the moment, leave the grade of supervenience unspecified.

evolves into a set of T-states that does not realize a U-state. An abstract example would be this. Suppose that β_1 is multiply realized by the set of T-states $\{\tau_1, \tau_2, \tau_3\}$. Suppose that the laws of temporal evolution in T are as follows: $\tau_1 \rightarrow \tau_1^*$, $\tau_2 \rightarrow \tau_2^*$, $\tau_3 \rightarrow \tau_3^*$ (thus we are assuming here T-temporal supervenience). We have TDD (so far as this example is concerned) if the set $\{\tau_1^*, \tau_2^*, \tau_3^*\}$ multiply realizes one U-state, which we would naturally label β_1^*. If, perchance, $\{\tau_1^*, \tau_2^*\}$ realizes one U-state while $\{\tau_3^*\}$ realizes another, TDD fails (since, for example, τ_1 cannot evolve into a realization of this latter U-state). Now, perhaps T-temporal supervenience fails. In that case instead of $\{\tau_1, \tau_2, \tau_3\}$ evolving to the determinate set $\{\tau_1^*, \tau_2^*, \tau_3^*\}$ we have an indeterminate evolution. For simplicity, let's confine the indeterminacy to τ_3 which can evolve either into τ_3^* or τ^{**}_3 (but τ_3^* and τ^{**}_3 cannot both obtain, no more than can any pair of possible realizers). Thus T-temporal evolution will lead from $\{\tau_1, \tau_2, \tau_3\}$ to $\{\tau_1^*, \tau_2^*, \tau_3^*, \tau^{**}_3\}$. Then TDD still holds if $\{\tau_1^*, \tau_2^*, \tau_3^*, \tau^{**}_3\}$ multiply realizes one U-state. If this set does not multiply realize a *single* U-state but rather underlies, say, two U-states, β_1^* and β_2^* where $\{\tau_1^*, \tau_2^*, \tau_3^*\}$ multiply realizes β_1^* and $\{\tau_2^*, \tau^{**}_3\}$ multiply realizes β_2^* then TDD fails even in the 'loose environment' where T-temporal supervenience does not hold (since, for example, τ_1 cannot evolve into a realization of β_2^*). But notice that it is possible to have TDD even if the set of T-realizers of some U-state do not evolve to a set which realizes a single successor U-state. This can occur if the T-realizers can each indeterministically evolve to a realizer of any permitted (by U) successor or the initial U-state, as in the figure 1 (see appendix for figures). [Go to figure 1]

Top-down discipline can exist from a supervening domain, U, to its supervenience base domain, T, even if T lacks T-temporal supervenience and U enjoys U-temporal supervenience. In such a case, we could say that there is 'de-randomization' of T (see figure 4 below). It is possible that the apparent deterministic character of classical (or macroscopic mechanics) is the result of this sort of de-randomization, as the quantum states that realize the macro-states provide top-down discipline for the macro-domain. That is, while there may be intrinsic randomness at the micro-level, it somehow cancels out in the myriad interactions involved in the realization of any macro-state.[8]

[8] A possible ground for de-randomization in the micro to macro relationship is given by Ehrenfest's equations, which assert that the *expectation value* of an observable such as position or momentum will evolve in accordance with classical laws of mechanics. In a macroscopic system made of huge numbers of microsystems we might expect

I think that much of the interest in multiple realizability which has been shown by philosophers lies in the possibility of a failure of top-down discipline rather than the mere possibility of multiple realization itself. For suppose that the supervenient domain, call it U, enjoyed top-down discipline with respect to its supervenience base, T. If there is T-temporal supervenience (and in physics we seem to have *full* temporal supervenience) then top-down discipline implies that there is also U-temporal supervenience (see R4 below). This would strongly suggest that the supervenient domain can be *reduced* to its supervenience base, since there is a set of subvening states that exactly map on to the theoretical relationships of the supervenient domain. That is, there would be a 'model' of U in the subvening T-states.

While we might expect that some domains, for example that of chemistry, enjoy both top-down discipline and a reductive relation relative to physics, this is not generally the case. Most domains 'above' physics, simply don't have (and don't necessarily even want to have) the resources to fully determine state-to-state transitions. If this lack of temporal supervenience in the supervening domain is coupled with the lack of top-down discipline (which I think is the usual case, since we presumably have physical temporal supervenience for the realizing states of the higher-level states), the case for reduction is very weak even though every supervenient state has, of course, a entirely definite, if very large, set of realizers in the supervenience base. This is because there is no model of the supervenient domain within the base. Consider figure 3 below. From the point of view of the T-domain situation, the U-state at t_1 has to be thought of as a disjunction of the particular T-states that can each realize that U-state. But that disjunction will not 'act like' a U-state since it transforms into a disjunction that cuts across U-classifications because none of the realizers act like the U-state in question. This seems to me a powerful reason for regarding supervenience without top-down discipline as a non-reductive relation between theories[9].

Notice however that it might be possible, in the face of a recognised failure of top-down

---

that such statistical features will exhibit a stability sufficient to allow us to identify the expectation value with the values obtained by particular observations, thus resulting in de-randomization and providing a reason to expect top-down discipline.

[9] There are other grounds for suspicion that such disjunctions of subvening states can support any robust sense of reduction, for which see Owens (19??) or Seager (19??).

discipline, to develop a more sophisticated, or discriminating, U-theory that differentiated the U-state at t_1 into two states that differed in some U-way so as to provide a U explanation of why the original (now regarded as *under-characterized*) state could evolve into distinct U-states. This new found difference within the U domain would reflect a partition of the set of T-realizers into two sets, each realizing a distinct U-state. This would result (so far as our diagram 3 goes) in top-down discipline. It would also make the case for reduction, as we would have succeeded in finding a set of realizers that 'act like' U-states throughout their dynamics. There might be some general theoretical pressure to search for such discriminators (see Sellars's discussion of the two kinds of gold, in Sellars 19??, and van Fraassen's comment in 19??). However, my own feeling is that most high-level theories are not in the business of giving such a complete description of their domain as to fully constrain its dynamics. The more fundamental a theory is taken to be the stronger this pressure will be; on the other hand, very high level theories, such as psychology, will hardly feel it at all[10].

If we take seriously, as we should, the possibility of indeterministic evolution of states, we ought also to consider that there may be definite probabilities associated with possible state transitions. We can easily adapt our notion of top-down discipline to incorporate this refinement.

Statistical Top-Down Discipline: A family of states (or theory), U, has *statistical* Top-Down Discipline (STDD, or U/T-STDD) relative to a family of states (or theory), T if and only if

(1) U supervenes upon T

(2) for every U-state, β, with permitted successors $\{\beta_1, \beta_2, ..., \beta_n\}$ and transition probabilities $\{p_1, p_2, ..., p_n\}$, each T-realizer of β, τ, has permitted successors $\{\tau_1, \tau_2, ..., \tau_n\}$ such that each τ_i can realize any of $\{\beta_1, \beta_2, ..., \beta_n\}$ and the transition probability of τ to $\tau_i = p_i$.

This is unfortunately complex but an example of a failure of statistical TDD should make it clearer. In figure 2 below [go to figure 2] we see a U-state that can indeterministically evolve to two U-states. The realizing T-states mirror the indeterminacy (they meet the initial definition of TDD given above) and manage to duplicate the correct U-transition probabilities overall (on the

[10] One good reason for this lack of concern is the recognition of distinctively lower-level 'intrusions' into the high-level dynamics which are simply not within the purview of the high-level theory. See Dennett (19??) for a classic discussion of this.

assumption that the U-state is equally likely to be realized by the two possible realizing T-states). But the T-transition probabilities do not mirror the U-transition probabilities at the level of the individual T-realizers. So statistical top-down discipline fails.

To my mind, the above situation is one that also counts *against* reduction of the U-states to the set of realizing T-states. These states do not behave like the U-states they realize at the level of individual states, though their interactions conspire to make the U-state description accurate at its own level. The particular probabilities in this case may seem somewhat miraculous (obviously they have been rigged to illustrate the desired failure of STDD), but the miracle is less pronounced in cases where high-level statistics emerge out of myriads of low-level processes and the high-level statistics reflect various 'laws of large numbers'. Furthermore, we expect that normally one theory precedes another and the follow-up theory must be constrained to produce known statistical relationships.

Normally, we take the possibility of indeterministic state evolution in the domain of a high-level theory to be the result merely of ignorance of – or sometimes unconcern about – underlying determining factors. Such factors may involve unknown, or imprecisely specified, features of the high-level theory or, more often I think, may involve features of lower level theories that 'intrude' upon the dynamics of the high-level theory to determine certain state transitions. For example, it seems entirely plausible to suppose that in certain cases the total psychological state of someone may not determine which action they will perform. Nonetheless, some action will be performed, and there may be no underlying psychological feature that accounts for this. There will, though, be some sub-psychological feature which tips the balance (in choosing between the cheese plate or the chocolate cake for dessert, perhaps the level of some neurotransmitter within some critical network of neurons plays a crucial role even though there is absolutely no conscious – or unconscious in a *psychological* sense of the term as opposed to a merely *non-conscious* – element of one's mental state that reflects that neurotransmitter level precisely enough to account for one's choice). If the underlying realizers of the high-level states are differentiated by features that do not correspond to differences in the high-level characterization, we would expect the probabilities of the state transitions of the realizers to differ from those of the high-level state transition probabilities and thus we would expect statistical top-down discipline to fail (in the limit

of temporal supervenience the probabilities of transition in the low-level theory go to one or zero). As noted above, this would also put a strain on a reductive account of the relation between the high-level states and the set of low-level realizing states.

§2. Relations.

The definitions given above result in a number of 'theorems' or relations between them that, in addition to their own interest, can be used to clarify a variety of possible views of emergence.

(R1) Final-physical-possibility does not imply physical-totality. The laws of final-physics may or may not be such as to sustain completeness, closure and resolution. It is, I think, the goal of many modern physicists to produce a theory that has totality. The structure of basic physics now in place would appear to be that of one aiming, so to speak, to be a total theory, but the current theory is manifestly incomplete (or worse, incoherent). We do not have a physical understanding of every physically possible situation, even at very basic levels. For example, we just do not know the physics of processes that essentially involve *both* quantum and gravitational processes (and the two fundamental theories involved, general relativity and quantum field theory may in fact be fundamentally inconsistent with one another). It does seem clear that the research on quantum gravity is designed to complete physics in a way that provides totality. But whether physicists can succeed in developing a final theory that is total depends not only upon their ingenuity but also upon the nature of the world itself. It is impossible to say now whether this research will or even *can* succeed for it is not given beforehand that there is a *single* physical theory that can encompass all elementary physical processes. Nor can we yet rule out the sort of radical emergence (to be defined below) that denies closure via resolution. Therefore we cannot say in advance that the final physics is a total theory or that the worlds that are finally-physically-possible are all such as to observe totality.

(R2) Strong supervenience of U upon T is compatible with the absence of T-temporal supervenience. Failure of T-temporal supervenience means that there is a T-state, α, that does not have a predecessor which leads uniquely to α. Obviously, this does not prevent U from strongly supervening upon T unless further conditions are met.

(R3a) Strong Supervenience of U upon T and T-temporal supervenience does *not* imply U-

temporal supervenience. The reason is that Top-Down Discipline of U relative to T might fail. That is, the set of realizers of some U-state(s) might lead to realizations of different subsequent U-states, even though each such realizer T-state has a unique outcome, as illustrated in figure 3 [go to figure 3].

(R3b) Nor does strong supervenience of U upon T and the *absence* of T-temporal supervenience imply the absence of U-temporal supervenience. This is possible because there could be top-down discipline of U relative to T despite the failure of T-temporal determination, as illustrated in figure 4 [go to figure4] (see also the discussion de-randomization above).

(R4) Strong Supervenience of U upon T, Strong T-temporal supervenience and top-down discipline of U relative to T implies Strong U-temporal supervenience.[11] If we have T-temporal supervenience then U/T-TDD implies that every T-state which realizes some U-state, β, must evolve to realize a single successor U-state. By strong supervenience , β must have a realizing T-state. So β must have a unique successor, which is to say, we have Strong U-temporal supervenience.

(R5) Strong supervenience of U upon T implies that U-states (probably) have T-efficacy. Suppose that U strongly supervenes upon T and consider some U-state, F. F has a set of T-realizers $\{\tau_1, \tau_2, ..., \tau_n\}$. To test if F has T-efficacy in the production of some state we consider the counterfactual:

> for some system, σ, and some actual outcome T-state, G of system π, if σ had not been in state F then π would not have been G.

In the nearest world where $\sim F\sigma$ we have $\sim\tau_1\sigma$ & $\sim\tau_2\sigma$ & ... & $\sim\tau_n\sigma$. We can assume that $G\pi$ was the outcome of one of $\tau_1\sigma$, $\tau_2\sigma$, ..., $\tau_n\sigma$ (since we need only find one such G to reveal efficacy). Since none of these obtain in the counterfactual situation, it is unlikely that $G\pi$ would come about nonetheless. Actual judgements of efficacy would have to depend upon particular circumstances, but it seems that it is very probable that states of strongly supervening domains have (or typically have) efficacy. To take a definite example, suppose that I want an apple and then reach out and take an apple. Was my desire efficacious in this transition? Well, according to strong

[11] Technically, supervenience is included in the definition of top-down discipline, but it is clearer to emphasize the role supervenience plays as a separate part of the proof. Also I did not specify the grade of supervenience in the definition of TDD but left it loose to implicitly form a family of relations of top-down discipline.

supervenience we assume that my wanting an apple was realized by some physical state, P, from the set of possible realizers of apple-wantings. In the counterfactual situation of my not wanting an apple, P – along with all the other possible apple-wanting realizing physical states – would not obtain (since if it did, by supervenience, my desire would exist after all). Would I still reach out for the apple? It is certainly possible to imagine situations where this occurs: suppose that Dr. Evil has trained his 'action-gun' upon me and is ready to force my body to reach for the apple at the appropriate time should I lack the desire. However, such situations of 'counterfactual over determination' are very rare, and thus we may conclude that strongly supervening states very probably – typically – have efficacy. (If T has *full* strong T-temporal supervenience then we can say that U-states definitely have T-efficacy. For then $G\pi$ would have a unique predecessor and if that predecessor did not occur then $G\pi$ would not occur. But then $\sim F\sigma$ would guarantee that the predecessor of $G\pi$ did not obtain and so $G\pi$ would not obtain. This may be of interest since physics appears to enjoy *full* strong temporal supervenience.[12])

(R6a) Strong T-temporal supervenience implies global supervenience for *any* domain with T-efficacy. Recall that to claim that U globally supervenes upon T is to say that any two worlds that agree on their assignment of T-states to systems will agree on their assignment of U-states. Symbolically,

$$(w)(w^*)(\sigma)(F \in U)((w =_T w^* \ \& \ F\sigma w) \supset F\sigma w^*).$$

Thus the denial of global supervenience would be expressed , after some manipulation, as

$$(\exists w)(\exists w^*)(\exists \sigma)(\exists F \in U)(w =_T w^* \ \& \ F\sigma w \ \& \sim F\sigma w^*).$$

That is, the denial of strong supervenience entails that there are indiscernible T-states that differ with respect to the non-supervening U-state, F. To test whether F has T-efficacy we must evaluate the counterfactual:

if $F\sigma$ had not been the case then $H\sigma$ would not have been the case,

where H is some outcome T-state which obtains in the 'source world' (i.e. the world from which

[12] Note that this argument does not lead to 'efficacy inflation' in the sense that the U-state in question helps to bring about every G-state in any system. My dream last night is not efficacious in producing an earthquake in Turkey even though that earthquake has a unique physical predecessor. On the assumption of full temporal supervenience, the nearest possible world in which I don't have my dream is different from the actual world right back to the beginning of time, but even so there is no reason to think it is different with regard to the earthquake's occurrence. In testing for efficacy, we can pick any outcome state we wish so we can find one for which my dream is efficacious. This does not lead to my dream's having efficacy 'everywhere'.

we will evaluate the counterfactual, w in the above) and which is putatively brought about by F. To perform the evaluation we consider the T-possible world most like the source world save for the differences necessitated by assuming ~Fσ. The T-possible world most like the initial world would be one that was identical with respect to T (up to the time of occurrence of Fσ), differing only with respect to F (and possibly other U-states). We know there is such a world, by the denial of global supervenience (w* in the above). However, by strong T-temporal supervenience, that world evolves over time in exactly the same way as the source world. Therefore the counterfactual is false and F cannot have T-efficacy, contrary to the assumption of R6a. So global supervenience must hold.

(R6b) Strong T-temporal supervenience implies strong supervenience for *any* domain with T-efficacy.

This argument is slightly less convincing than that for R6a, because we need an additional assumption. Suppose we have T-temporal supervenience but there is a T-efficacious domain, U, that does not strongly supervene upon T. Then by the definition of strong supervenience, there is a T-possible world where there is a system, σ, and U-state, F, such that

$$F\sigma \;\&\; \sim(\exists G \in T)(G\sigma \;\&\; \Box(s)(Gs \supset Fs))$$

So we have Fσ and

$$(G)(\sim G\sigma \vee \sim\Box(s)(Gs \supset Fs))$$

Now, this means either (1) that σ is in no T-state whatsoever or (2) there is such a state but it does not necessitate F. If the former, then σ is a radically non-T system. Suppose F has T-efficacy. Then the presence of F makes a difference in a T-state. But since F is a state characterizing utterly non-T entities, the presence or absence of F is not marked by *any* necessary T difference. For while it is perhaps possible to imagine that there might be some kind of a 'metaphysical' connection between some T-state and the presence of F, this connection is not a T-law (T-laws do not say anything about radically non-T objects). Violation of this connection is thus not a violation of any T-law, and the world in which this connection is broken is thus a T-possible world. So, given T-efficacy, there could be two T-indiscernible situations which differed in their outcome because of the difference in F. But this violates strong T-temporal supervenience. That is, since F is not marked by any T-state we can take the F world and the ~F world to be T-

indiscernible (and worlds can't get any more similar in T-respects than *T-indiscernibility*), then use the argument for R6a. Now, suppose that strong supervenience fails because of (2). Then there is a T-state, G, that σ has but is such that G does not necessitate F. This entails that there is a world in which some system has G but does not have F. We might then try to argue that in *every* world, G has the same outcome by strong T-temporal supervenience. Thus in whatever world we choose to evaluate the counterfactual which tests for the T-efficacy of F, there will be no T-difference. Therefore F does not have T-efficacy – it cannot make any difference. But this won't quite work as it stands since it is open to the following worry. The counterfactual test requires that we go to the world most similar to the source world save that ~Fσ holds. What if this is a world where ~Gσ holds? Abstractly speaking, this seems to be possible. However, such a world will be quite unlike the source world, since strong T-temporal supervenience requires that G's predecessor not appear in the test world (else we would get G after all) or else we have a miracle (which immediately violates T-temporal supervenience). That is, the assumption of ~Gσ propagates other T-changes throughout that world. Thus it is very plausible that a ~Gσ world is not the most T-similar to the source world and, after all, we *know* that there is a world in which Gσ and ~Fσ. If this is correct then the test world contains Gσ and hence must evolve to the same successor state as the source world, thus revealing that F does not possess T-efficacy[13].
Since strong supervenience implies weak supervenience we trivially get Strong T-temporal supervenience implies weak supervenience of T-efficacious domains. It is also the case that since strong supervenience implies global supervenience we have R6b implies R6a. Furthermore, since strong supervenience implies what I called local supervenience, we also get that strong T-temporal supervenience implies local supervenience.

Note also that we have to assume T-efficacy in the above since nothing can rule out the possibility that there are 'parallel' domains that do not supervene upon T but rather exist entirely independent of the T-world yet enjoy rich causal relations amongst themselves, a situation that would be approximated by considering one of Leibniz's monads *without* the pre-established harmony. The assumption of T-efficacy forges an essential link between the U and T domains.

[13] Note we must assume *strong* T-temporal supervenience to get this result, since in considering strong supervenience we have to consider other physically possible worlds.

Such an assumption is reasonable since we have little interest in hypothetical domains that are entirely isolated from each other. In particular, we are not very interested in an epiphenomenalist view of the mind-body relation, though it is important to see that epiphenomenalism cannot be ruled out by any considerations advanced thus far. It is also interesting to note that, given (R5), we have it that strong T-temporal supervenience implies that U is T-efficacious if and only if U strongly supervenes upon T.

This highly interesting and perhaps initially surprising result reveals the significance of temporal evolution of states for the metaphysics of dependence. If we have a domain the states of which evolve through time according to the laws of that domain, then there are tight constraints placed upon the states of any other domain which are to have effects within that initial domain. They must 'ride upon' the lawful transitions of the initial domain to both preserve those lawful transitions and have their own efficacy, which is to say, they must supervene upon the states of the initial domain.

(R6c) *Weak* T-temporal supervenience implies weak supervenience for any domain with T-efficacy. The argument for this claim is still weaker since additional assumptions (or modal intuitions) are needed. The argument proceeds almost as that for R6b. But when we consider the first horn of the dilemma, that σ might be a radically non-T system, we must consider the counterfactual, if σ had not been F then things would have been T-different. It seems to me that the closest world in which σ is not F is one in which the T-temporal supervenience relations are not altered (since F has nothing whatsoever to do with T, it is hard to see why the T relations would be different in that world). If so, F's T-efficacy would fail. (The alternative idea, I guess, is that because of some kind of pre-established harmony, in the nearest world where σ is not F, the T-temporal supervenience relations must be altered enough to make the counterfactual come out true. But even in such a case, it seems that it is the alteration in T that accounts for the difference in outcome so that intuitively F has no efficacy in the T domain after all.) The other horn of the dilemma leads to the claim that there is an object, π, in the very same world as that in which σ has F such that π has G but does not have F. Then in that very world we have a test of F's efficacy and – because of weak T-temporal supervenience – within any world the T-temporal supervenience relations are the same. Thus G will lead to the same outcome for π as for σ. So F's

T-efficacy seems to fail. If it is insisted that 'singular causation' is possible then we must use the counterfactual test, and then we can employ the plausibility argument given above.

(R7) T-Totality implies strong T-temporal supervenience (up to intrinsic randomness of T). Totality is a very strong condition on the nature of the laws of a theory as well as on the 'metaphysical structure' of the world (roughly, constituent structure with 'bottom-up causation' sufficient to yield all phenomena). But is it enough to guarantee temporal supervenience? Let us see. Assume that T is (supposed to be) a total theory but that T-temporal supervenience fails. Then there is a T-state, G, of system σ that does not have a unique outcome (let's say that in such a case Gσ *diverges*). If Gσ is a complex state then by the property of totality I labelled 'resolution' we can resolve it into a set of elementary T-constituents that act entirely according to T-laws. If Gσ does not have a unique outcome this must be because some elementary state does not have a unique outcome. So we might as well consider Gσ to be such an elementary state. It is impossible for Gσ to diverge because there is a sub-T theory which realizes the T-states and which accounts for the divergence of Gσ. For then not everything that happens would be the result of the operation of T-laws and T-totality would be violated. The only possibility of divergence is if T has some intrinsically random elements within it. That is, if it is a brute fact that for some T-state two (or more) distinct states can ensue. For example, on certain views of quantum mechanics (e.g. those that espouse the 'uncontrollable collapse of the wave function' upon measurement) QM-temporal supervenience fails. A particular uranium atom, in state G, may or may not fission. If it does we get, say, state G_1; if it does not we get state G_2. There is nothing within quantum mechanics to account for this (and no hidden variable lurking beneath quantum mechanics either). The fissioning or lack of fissioning at any particular time is intrinsically random. If there is no intrinsic randomness then it seems that totality implies temporal supervenience. We could leave this result there: if there is no intrinsic randomness in the elementary states of T then totality implies temporal supervenience (this is less trivial than it appears since high-level theories can fail to observe temporal supervenience without possessing intrinsic randomness; totality implies that the lack of temporal supervenience must result from *intrinsic* randomness, not the sorts of intrusions from below that characterize high-level theories). In fact, it implies strong temporal supervenience since totality is a property of the laws of a theory and so naturally sets the

conditions of possibility relative to that theory.

However, there is more to say about intrinsic randomness. It is important to see that the possible existence of intrinsic randomness does *not* fundamentally change our result. To take account of this possibility we would have to complicate our definitions considerably, along the following lines. In place of individual states we would have to take probabilistically weighted *sets* of states. We could then recast our arguments in these terms. Instead of a unique outcome state as the defining characteristic of temporal supervenience we would have a uniquely statistically weighted set of states. Although this would get very messy I think in the end we would get completely analogous results to those obtained when we do not consider intrinsic randomness. A form of *statistical* temporal supervenience would be defined in terms of predictably weighted ensembles of states

As an illustration, consider a view once defended by John Eccles (see ??). In support of a form of Cartesian dualism, Eccles hypothesized that perhaps the mind could act under the cloak of quantum mechanical indeterminacy, subtly skewing the intrinsically random processes occurring at the synapses of the neurons. This is conceivable, but it would be experimentally revealed, in principle, by noting that the distribution of outcome states of synaptic conditions did not strictly observe the statistics predicted purely on the basis of quantum mechanics. In this way, quantum mechanics would be refuted. If quantum mechanics is true, then the mind can only act in accordance with the statistics predicted by quantum mechanics – and this would bear out the statistical version of totality. This reveals that intrinsic randomness within a theory only complicates temporal supervenience but does not destroy its essence.

We could define a state's, F's, *statistical* efficacy within a theory that allows for some intrinsic randomness as the presence of F making a difference to the outcome statistics over repeated 'counterfactual trials'. For example, adding some weight to one side of a die is statistically efficacious for while it does not prevent any number from coming up it does change the outcome statistics over many trials (perhaps only very subtly).

(R8) Strong Supervenience of every T-efficacious domain, U, upon T and strong T-temporal supervenience implies T-Totality. Suppose every T-efficacious domain, U, strongly supervenes on T but that T-totality fails. Then either closure, completeness or resolution fails. If

completeness fails then there is an entity which has no (non trivial) T-description, a radically non-T object. This entity must be from some domain, U. But then there *could be* a difference in U with no difference in T, for while it is perhaps possible to imagine that there might be some kind of a 'metaphysical' connection between T-states and the U-states, this connection is not a T-law if U is a radically non-T domain. Violation of this metaphysical connection is thus not a violation of any T-law, and the world in which this connection is broken is thus a T-possible world. But this violates strong supervenience. Suppose, then, that closure fails. Then for some domain, U (which, here and below, might be T itself), some U-state, β, occurs in violation of some T-laws (say that β is a *miracle*). But – by strong supervenience – β has a realizing T-state, τ[14]. By strong T-temporal supervenience, τ has a predecessor state, π, for which τ is the necessary unique outcome. Could τ occur but occur in *violation* of T-laws? No, for then it would be T-possible for τ not to occur even though its predecessor state does occur. If it is not a matter of T-law that π led to τ then there is a T-possible world where we have π and $\sim\tau$. But that violates T-temporal supervenience. Therefore, τ's occurrence is not in violation of any T-law. Since τ is the realization of β, β's occurrence does not after all violate any T-law, so closure cannot fail. Finally, suppose that resolution fails. Then there is a domain, U, and a U-state, β, such that either there is no description of β in T-elementary terms or there is such a description but the presence of a particular instance of β leads to system behaviour distinct from the behaviour of β's elementary T-constituents as they would act under the T-laws governing the elementary T-constituents (let's label this possibility the divergence of β's behaviour from that of β's elementary realizers – the shadow of emergence is obviously looming here). The first disjunct violates completeness[15]. On the second disjunct, there must be a T-state that subvenes β, call it τ

[14] Or, more strictly speaking, a set of possible T-realizers $\{\tau_1, \tau_{2, \ldots,} \tau_n\}$. The argument is not affected by this detail, which is thus omitted for simplicity.

[15] Here I assume that if there is a T-description of a system then there is a description in T-elementary terms. This is an innocuous assumption since, by itself, it does not imply that every T-state has a *constituent* structure formed out of T-elementary features, for maybe some 'large' T-states are themselves elementary. It is hard to think of genuine examples, but here is a possibility. Blackholes can have but three physical properties that fully characterize them: mass, charge and angular momentum. These properties are a function of the properties of the elementary constituents that have formed the blackhole. But, once formed, there is no sense in which the blackhole is *composed* of little bits and pieces that individually have various masses, charges or angular momenta (string theory may alter our perspective on this, but, of course and interestingly, in a way that makes blackholes resolvable into a new – but still physical of course – kind of elementary constituent structure). Thus the blackhole cannot be resolved

which is composed of a set of elementary T-features $\{\tau_1, \tau_2, \tau_3\}$ (we know we have this decomposition by way of the assumption that resolution fails via divergence). T-temporal supervenience means that there is a unique outcome of each τ_i, so $\{\tau_1, \tau_2, \tau_3\}$ has a unique set of elementary T-features as its outcome. Therefore, divergence of β's behaviour from that of β's elementary realizers violates T-temporal supervenience[16]. Since we assume that T-temporal supervenience holds, such a β cannot exist, and therefore resolution holds. So T-Totality follows.

(R8b) Global Supervenience of every T-efficacious domain, U, upon T and strong T-temporal supervenience implies T-Totality. The argument proceeds exactly as for R8a with respect to completeness. Suppose, next, that closure fails. Then for some domain, U, some U-state, β, occurs in violation of some T-laws. Now, for global as opposed to strong supervenience the idea of a realizing state is more vague, but there must be a state of the *world* – call it τ – such that any world T-indiscernible to τ up to the time when β occurs will agree on all U-assignments and thus is a world in which β obtains. According to strong T-temporal supervenience, a world T-state restricted to time prior to β's obtaining is sufficient to guarantee that β will obtain. We might label such a temporally restricted world state $\tau_{<t}$. Any two worlds that are in state $\tau_{<t}$ will end up agreeing about β. Strong T-temporal supervenience requires that $\tau_{<t}$ lead to the world as it is at the time of β's obtaining (call that state $\tau_{<=t}$). The process from $\tau_{<t}$ to $\tau_{<=t}$ cannot violate any T-law (for the same argument given in the proof of R8a), so β's coming to be cannot violate any T-law. So we have closure. Finally, if we suppose that resolution fails then either completeness will fail as in the proof of R8a or we have divergence of β's behaviour from that of its elementary realizers (which might, under global supervenience extend to the elementary state of the entire world). But such divergence would violate T-temporal supervenience (again as above, using our machinery of temporally restricted world states). So T-totality follows.

(R9) Strong T-temporal supervenience implies T-Totality (across domains with T-efficacy).

---

into sub-components. This is no violation of the totality of physics however, since charge, mass and angular momentum are themselves allowable elementary features. A blackhole is, so to speak, a kind of elementary 'particle' (and one that can, of course, take a constituent place within larger physical assemblies such as multi-star systems, galaxies, etc.).

[16] Notice we do not need to assume that U possesses top-down discipline for this argument to work. The single case of β's divergence violates T-temporal supervenience.

From above (R6a) or (R6b), strong T-temporal supervenience implies Strong T/U supervenience or global T/U supervenience for any domain with T-efficacy. Therefore, from (R8a) or (R8b) the result follows.

(R10) Strong T-temporal supervenience if and only if T-Totality (across domains with T-efficacy). Various forms of this follow from (R9) and (R6).

§3. Emergence.

Emergentism is the doctrine that certain features of the world – features of the emergent domain – *emerge out of* other features from another domain, call it the *submergent* domain. To say exactly what 'emergence' is and how it works, is not so easy. The simplest view, and one that dovetails with the approach of this paper, is to regard emergence as relative to theoretical descriptions of the world. A feature is emergent only if it is part of one theoretical description but not another. For example, the valence of an atom is emergent inasmuch as it forms a part of chemical theory but not a part of physical theory (i.e. physics). Or again, the 'fitness' of a genome is an emergent feature insofar as it is utilized by evolutionary biology but not, for example, by chemistry.

Of course, this criterion is but a part of what it is for a feature to be an *emergent* feature. We must add a notion of the 'direction' of emergence, for while valence is a good example of an emergent feature we are not inclined to call *spin* an emergent just because spin is not mentioned in evolutionary biology. The 'direction' of emergence brings supervenience into the picture in a natural way. For the additional idea is that of *determination* of the emergent feature by features of the submergent domain. Thus, we find it appropriate to say that valence is determined by physical features, but have no reason at all to suggest that spin is determined by features peculiar to evolutionary biology. It is the nature of this determination that clouds the issue of emergentism, and suggests that work on supervenience may be of assistance in its clarification.

For example, if we have strong supervenience of U upon T then we have what are in effect 'laws of emergence' that are constant across all T-possible worlds. These laws of emergence are expressed in the latter part of the formula definition of strong supervenience (i.e. the '$\Box(\sigma)(G\sigma \supset F\sigma)$' (where, recall, $G \in T$ and $F \in U$) part of the definition). And this is another reason for

preferring strong supervenience over global or local supervenience – it finds a definite T-state as the base for the emergent properties and this is in line with most emergentist thought[17]. If we consider the difference between strong and weak supervenience in terms of emergence, we see that weak supervenience allows for the laws of emergence to vary across submergently possible worlds, which is an interesting and, as we shall see, actually critical component of any serious form of emergentism.

One digression. Certain properties can perhaps be called emergent even though they fail to meet our first criterion. Mass, for example, figures in physics, yet the mass of a physically complex object can be thought of as an emergent. This is a 'mereological' sense of emergence, roughly characterized as a feature which an object has but which no proper part of the object possesses, although the parts possess 'cognate' properties. Thus, 'having a mass of 1 amu' is a property of an (ordinary) hydrogen atom, but none of its proper parts have this property. This seems to me rather a degenerate sort of emergence, for the 'generic' property -- the determinable if you will, in this case 'mass', equally applies to both the whole and its proper parts. It is to be expected that a supervenience relation also holds between the submergent properties and the mereologically emergent properties, and usually one that is pretty straightforward and unlikely to lead to any substantial issues of emergentism.

In marking out the central features of emergentism we must begin by contrasting emergentism with dualism. Emergentism is anti-dualist; emergent features are features of objects which always have descriptions – albeit incomplete insofar as they neglect the emergents – from within the submergent domain. Emergence does not generate a realm separate and apart from the submergent domain. A second crucial feature of emergentism is the denial of epiphenomenalism;

[17] However, this at least suggests that there may be novel emergentist doctrines that derive from global or local supervenience relations. Perhaps we can imagine emergent properties that depend upon total world states for their existence. These are emergent properties dependent upon the total state of the 'whole universe' even though they might be properties of individual things. I can't think of any examples of such properties however, although there are clear cases of non-local emergents. 'Being money' is such a non-local (but hardly fully global) emergent, but because of its lack of efficacy and our possession of some idea of how we might explicate the existence of money in terms of non-monetary properties, we regard this as a form of benign emergence. Another example of a very non-local but far from fully global emergent property might be the value of the gravitational field at any point; it may well be that the state of the entire universe figures in determining this value (though perhaps not, depending on whether there are regions of the universe that are not in causal contact, which currently seems very likely). The important point made by these examples is that even in non-local emergence, the emergent property depends upon quite definite, if 'spread out' features of the submergent domain.

emergent properties are supposed to be efficacious, their presence makes a difference to the way the world goes. However, the nature of this efficacy is not always clear and can vary from a weak to a very strong claim about the role of the emergents in the unfolding of the world.

We can use the results of the previous section to define the two fundamental types of *emergence* (along with an odd and probably useless additional variant). The weakest form of emergence is one which offers no threat to the operation of the submergent domain from which the emergents spring. To put it another way, the existence of such emergents is explicable (in principle, as discussed below) on the basis of the submergent domain. Examples of such emergence are, presumably, the liquidity of water, the shape of macroscopic objects, the chemical properties of substances, the weather, etc. Such an emergence poses no dualist threat – the emergents are clearly features of systems describable in submergent terms. And emergents of this kind can be said to have efficacy. The view that meets these conditions is what I'll call *benign* emergence.

U Benignly emerges from T if and only if T is a total theory and U has T-efficacy. If T is a total theory, then U strongly supervenes upon T (if U has T-efficacy), so that we have an explication of the origin of emergent properties based upon the elementary T-features into which every U feature can be resolved. Such emergents can have efficacy, in the sense that complexes of elementary T-features can have efficacy. That is, it seems easy for such emergents to pass the counterfactual test of efficacy, and hence they will meet the definition of efficacy given and used above. Nonetheless, everything that happens, including the combinations of T-elementary features that underlie the emergents, happens in accord with the laws of T.

It is worth pointing out that when I say that under benign emergence we would have an explication of emergence in terms of the submergent domain I do not mean that the explication would be simple. It might be of such complexity that it will remain forever beyond our full comprehension. Generally speaking, these explications will proceed on a case by case basis, by the deduction from T-states and T-laws of all the behavioural capacities of U-states as well as the deduction of U-laws as springing from these behavioural capacities. We already know enough about complex systems to be quite sure that the detailed explanation of many emergents will be beyond our best efforts.

A recent example illustrates both the nature of benign emergence and the need for an 'in principle' clause (I draw the example from DiSalvo 1999). We've known for a long time how to perform thermoelectric cooling – the effect was discovered in 1834 by Jean Peltier (you can now buy specialty picnic coolers that operate thermoelectrically). The advantages of such cooling include compact size, silent operation and no moving parts, but applications have been limited by the low efficiency of current materials. Thermoelectric cooling operates at the junction of two different conductors, one containing positive charge carriers (called holes), the other negative charge carriers (electrons). Passing a current through the junction causes both sorts of charge conductors to move away from the junction, thus carrying heat away from the junction. While this is a extremely over simplified and highly schematic explanation, it reveals how thermoelectric cooling is benignly emergent. The efficiency of the process is critically dependent upon the nature of the conductors forming the junction however, and is expressed in a parameter known as ZT. Known materials have a ZT of about 1; if materials of ZT around or above 4 could be found, thermoelectric cooling would vie with conventional methods of refrigeration for efficiency. Unfortunately, there is no general and practical way to accurately predict the ZT of a substance. DiSalvo explains the situation thus:

> Understanding electrical carriers in crystalline solids is one of the triumphs of modern quantum mechanics, and a theory of TE [thermoelectric] semiconductors has been available for about 40 years. This transport theory needs one input: the electronic band structure. More recent advances in determining the band structure, based on density functional theory and modern computers, give acceptable results. The main input to band theory is the crystal structure of the material. Known compounds can be sorted into a much smaller group of crystal structure types. A given structure type may be adopted by many compounds, and by comparison, we can often predict which elemental compositions will have this same structure because of similar atom sizes and average valence, for example. However, many new ternary and quaternary compounds adopt new structure types which cannot be predicted beforehand, and without the crystal structure, electronic band structure cannot be calculated. Not only is the inability to predict crystal structure (and thus composition or properties) the main impediment to predicting which new

> materials will make better TE devices, this inability is most often the limiting factor in obtaining improvements in most other materials applications. (1999, p. ??)

This inability to predict benignly emergent properties stems from a failure of our grasp of theory and/or our inability to perform extremely complex calculations. There is no real question that a 'mathematical archangel' – to use Morgan's evocative term – unfettered by limitations of computational speed or memory capacity would deduce ZT from quantum mechanical principles and the basic physical structure of the candidate materials.

More abstractly, if we have a total T-theory then we can in principle explicate the behaviour of any system of any complexity from a knowledge of its elementary T-structure. We know from totality, that all systems have such a structure and closure guarantees that such predictions are in principle possible (they may, of course, yield only statistical results depending upon the nature of the T-theory).

So, benign emergence is the model of emergence one must adopt if one accepts that physics is (or will be) a total theory. And most philosophers do attempt to take this route (see Kim, Searle, etc.). It may be the natural view of emergence from within the 'scientific view of the world', since that view is taken by very many thinkers to include the claim that the world is total (that is, that physics, which provides the fundamental description of the world, is a total theory). But I would like to remind the reader that, as discussed above in (R1), no one knows if the final physics will be a total theory, and hence no one knows if the fundamental structure of the world is total either. Whether or not the world is total is an empirical matter, and cannot be decided by any a priori metaphysical arguments.

The original emergentists, which include Mill, Lewes, Morgan, Alexander and Broad, would not have been satisfied with mere benign emergence (for an excellent general discussion of their views, see McLaughlin ??). They wanted more, and in particular they wanted their emergents to possess both a stronger form of efficacy and a more 'mysterious' relation to the submergent domain than benign emergence allows. Furthermore, although the move from submergent to emergent was to be mysterious it was to be a part of the natural order, not a mere accident or lucky chance. That is, the presence of an emergent feature was supposed to be in principle unpredictable even given a completely precise specification of the submergent domain and a

complete theoretical understanding of it. A sign of emergence is, as C. D. Broad put it, '... that in no case could the behaviour of a whole composed of certain constituents be predicted merely from a knowledge of the properties of these constituents, taken separately, and of their proportions and arrangements in the particular complex under consideration' (MPN, ??).

The point of talking of 'prediction in principle' is to provide a natural way to erase the epistemological constraints which can cloud metaphysics. The claim of impossibility of prediction of U-states on the basis of fundamental T-state even in principle is the denial of *determination* or strong supervenience of U upon T. It is conceivable that this venerable way to approach the ever present gap between epistemology and metaphysics which links in principle predictability with strong supervenience masks another distinction, a distinction between predictability (in any sense) and determination. If so, the deployment of the idea of prediction in principle would become (even more of) a *metaphor* for the determination of all properties but those of the submergent domain. But I take it that Broad and the other emergentists did intend to speak of a lack of determination or supervenience when they talked of a lack of predictability in principle and I will follow them in this.

Let us call this hypothetical, new form of emergence *radical* emergence. It is obvious that radical emergence implies that the submergent domain is *not* total (or that the theory of the submergent domain is not total). The failure of totality can be further diagnosed as a failure of closure. Completeness can hold, since the emergents are not new substances; and resolution can hold in the sense that complexes that possess emergent properties can be resolved into elementary constituents of the submergent domain. But the behaviour of these complexes is – most emphatically – *not* given by the concerted behaviour of those elementary constituents as they act, or would act, solely under the laws of the submergent domain. Thus closure must fail. We know from R10 that the failure of totality implies that we do not have strong T-temporal supervenience.

So if radical emergence is true than physics is not total. This could obtain in two ways. The first is that physics, as a theory, could be merely formally total. That is, physics could have the form of a total theory but be *false* of the world. Right now, given the pretensions of physics and its structure, this seems to be the only way radical emergence could be true. It is from this viewpoint that a severe tension is generated between radical emergence and physical theory. But

the other way totality can fail is, I think, more promising. It is possible to imagine physics just giving up its efforts to be total and resting content with describing the nature of the 'ultimate constituents' of the world with no implication that this description will implicitly fully constrain all of the world's 'behavioural possibilities'. It will, that is, be possible to resolve every complex physical entity into ultimate physical constituents, but not possible, even 'in principle', (and not thought to be possible) to recover the behaviour of every such complex merely from the interactions of the constituents as they act according to the laws of fundamental physics.

This is indeed a *radical* departure from our usual understanding of the aim of physical theory, for it requires a physics that is essentially 'uncompleteable', one admitting that the transition from elementary physical activity to the activity of complex physical systems is not entirely governed by fundamental physical law. Thus it is implausible to modern sensibilities. And this implausibility may be grounded in more than emergentism's unfashionable opposition to the physicalist *zeitgeist*, since emergentism may contradict some of the very general principles upon which our modern physical understanding of the world is based. But it is difficult to decide whether radical emergence actually *requires* the violation of such principles. For example, does radical emergence entail the violation of the principle of the conservation of energy? It seems that it might not, and there are at least three ways to arrive at this conclusion. However, one of these ways, due to McLaughlin (19??), reveals the almost irresistible urge back towards the totality of physical theory and the consequent demotion of radical emergence to mere benign emergence.

McLaughlin's suggestion is that where we have a system with emergent features acting in a way that appears to diverge from the action we would expect based on the physical understanding of the constituents of the system, thus violating the conservation of energy, we can reclaim energy conservation by positing a new sort of potential energy field which the emergent features can, so to speak, tap. The difficulty with this solution is that this potential energy field will naturally be counted as a new and basic physical feature of the world, which restores totality to physics and with it predictability (in principle) of the behaviour of complex systems from a knowledge limited to all the fundamental physical features of the system in question.

An example to illustrate this problem is the famous Casimir effect, which at first sight may seem to offer an instance of radical emergence. If two flat metal plates are placed very close to

each other (but not touching) there will arise a force between them, pushing them together ever so slightly. Is this the radical emergence of a new force emerging from certain quite particular macroscopic configurations of matter? No. The explanation of the Casimir effect is, roughly, that there is an energy in the vacuum which is, because of the nature of the metal plates and arcane details about the possible distributions of virtual photons between and beyond the plates, slightly greater outside the plates than between them. The point here is that the explanation spoils the appearance of radical emergence, for the 'potential energy' locked in the vacuum is explicable in elementary terms, and thus the force between the plates is predictable (in principle) just from basic physical features.

McLaughlin's proposal, then, is a general method of transforming radical into benign emergence, by the postulation of new potential energy fields which can be regarded either as stemming from or as themselves constituting new elementary physical features of the world. That is, these fields might be explicable in more elementary physical terms (rather as in the example of the Casimir effect) or they might be new brute facts about the physical structure of the world.

The second proposal retains the radical-ness of emergence but requires that there be a high-level 'conspiracy' to balance the energy books. That is, the defender of radical emergence can believe that energy is created 'out of the blue' when certain complex configurations are realized but that somehow an equal amount of energy disappears from the universe 'somewhere else' whenever these configurations arise. This is not impossible, but would of course be utterly mysterious from the point of view of basic physics.

With respect to this 'defence' of energy conservation, it seems the defender of radical emergence might do better to simply allow the conservation of energy to lapse on the grounds that it is better to have one mystery rather than two (the second being the odd and oddly coordinated *disappearance* of energy). After all, if we are allowing radical emergence there is no reason to deny that *energy* can radically emerge.

But a third method of saving energy conservation is perhaps more in line with radical emergentism and its assertion that fundamental physics is uncompleteable. The main idea here is that energy conservation is a *system relative* property, and those systems exhibiting emergent properties will abide by energy conservation as systems, with no implications about the processes

involved in the system's coming into being. What I mean can best be explained by an example. Physical systems can often be described mathematically in terms of a certain function, called the Hamiltonian, which encodes relevant properties of the system as well as forces acting on the system. The simplest case of use to us is the classical Hamiltonian of a single particle constrained to move in a single dimension subject to a field of force. The mathematical expression is this:

$$H(x, p) = \frac{p^2}{2m} + V(x)$$

Here, *p* represents momentum, *m* mass and the function *V(x)* represents the 'force field' in which the particle moves. From this equation one can deduce the functions governing the position and momentum of the particle over time. Most significantly, the Hamiltonian function is an expression of the energy of the system, and it can be shown that the time rate of change of *H(x,p)* is exactly 0, i.e. that the energy of the system cannot change over time. But notice that this description of our system says nothing about the nature of the 'particle' involved, and nothing about the nature of the force which governs its motion. So a system with emergent properties could instantiate this description *at the level* of the emergent features. The radical emergentist regards as another matter altogether the issue of whether, or how, the constituents (entities or processes) of this system unite or combine to create the whole system. Thus we are free to regard energy conservation as a constraint only upon systems as such. For if radical emergentism is true, there is no way to understand the creation of complex systems entirely in fundamental physical terms. Simple, non-emergent systems will obey the principle of the conservation of energy and so too will complex systems with emergent properties. The transition from simple, non-emergent to complex, emergent systems is not explicable by basic physics and is thus not bound by principles restricted to fundamental physics.

Although radical emergence denies the totality of the submergent domain, it is an open question whether we could allow strong supervenience within our radical emergence, since while non-T-totality implies non-T-temporal supervenience, it does not imply that strong supervenience fails.

However, it is easy to show that strong supervenience should not be accepted within radical emergence, for it would make the emergent features objectionably unexplanatory or, in a

way, epiphenomenal. For consider that the lack of T-temporal supervenience means, at least, that it is possible for two indiscernible T-states to have different outcomes. If these T-states are the base for an emergent property then, if we allow strong supervenience, then they will subvene the same emergent property. Therefore the emergent property will be unable to explain why there is divergence when you have T-indiscernible initial states. The lack of T-temporal supervenience is 'brute' (relative to U at least). If we want T-divergence to be explained by the emergent features then we cannot have strong supervenience.

To make this argument slightly differently, classical emergentists believed that the behaviour of complexes was in principle unpredictable from a knowledge – however complete – of the states and relations of and the laws governing the elementary submergent features. They nonetheless took it that the emergents were supposed to *explain* the divergence of the behaviour of the complex from the behaviour of the complex as it would be if it were determined solely on the basis of submergent laws and states alone. But as noted, if we have strong supervenience then the complexes would always subvene the same emergent feature (if any). If the behaviour of the complex was the same in all possible worlds then we would recover temporal supervenience and hence totality and we would be back to benign emergence. The action of the complex would after all be predictable on the basis of the state of the elementary submergent features constituting the complex[18]. Thus if the emergents are to explain the divergent behaviour of complexes, we cannot have strong supervenience.

Although mysterious from the point of view of fundamental physics, the emergentists thought that the emergence of high-level features was nonetheless a part of the natural order. Once we know that, for example, a particular chemical property arises from the combination of certain basic physical entities, we can infer that this chemical property will arise whenever we have this physical combination. As Broad puts it: 'No doubt the properties of silver-chloride are completely determined by those of silver and of chlorine; in the sense that whenever you have a whole composed of these two elements in certain proportions and relations you have something with the characteristic properties of silver-chloride' (MPN, p. ??). But this relation is 'a *law* which

[18] And note even if we have a lack of T-temporal supervenience due to intrinsic randomness in T, strong supervenience rules out the emergents explaining it – we are back to 'statistical totality'. (See the discussion of Eccles's manoeuver above.)

could have been discovered only by studying samples of silver-chloride itself, and which can be extended inductively only to other samples of the same substance' (MPN, p. ??, my emphasis).

Thus it seems that since the emergence is not a product of T-laws acting by themselves, there are T-worlds that differ with respect to the emergents attendant upon the same T-state (this is a variation that does not violate any T-laws). But at the same time, within these worlds we can generalize the emergence of emergent features across all *intra-world* instances of indiscernible T-states. This is an exact statement of a claim of *weak supervenience* of the emergent features upon T-features. We can illustrate the situation in figure 5 [go to figure 5). And we can define radical emergence as follows:

Radical Emergence of U from T = weak supervenience of U upon T + non-totality of T.

There is an interesting and perhaps rather attractive symmetry of causation in radical emergence that is lacking in doctrines that espouse totality, such as benign emergence. The failure of totality under radical emergence is explicable in terms of a very strong form of 'top-down' causation. Totality will fail when complex systems fail to act in the ways they should act if their behaviour was entirely generated by the interactions of their constituents according to the fundamental laws governing those constituents and their interactions. Our label for such failure is *divergence*, so we can say, in short, that divergence is explicable by top-down causation. Now, as noted above, the divergence of complex systems as described by *high-level* theories is commonplace, and such divergence is explicable by bottom-up causation; we expect that high-level generalizations will fail because of intrusions 'from below' of effects stemming from lower-level processes or structures. Radical emergence accepts that there will be exactly analogous, and genuine as opposed to the merely apparent – or at least entirely explicable in low-level terms – top-down causation found in total theories, intrusions 'from above' as well, so that complex systems, described in terms of low-level theory, will suffer from effects stemming from higher-level processes or structures, effects which are not predictable solely from the low-level state of the systems.

Another interesting feature of radical emergence is that it tends to conspire to give an illusion of totality. That is, radical emergence of U from T entails weak T-temporal supervenience (up to intrinsic randomness of T). Thus, within a world, T-complexes that are indiscernible all act exactly the same (or, at least, generate the same behavioural statistics). Such a world could 'look'

like it was T-total and encourage the search for a total T-theory. A rather bizarre further 'metaphysical' possibility is that such a 'total' theory could perhaps, given sufficient ingenuity, be found despite the existence of genuine radical emergence. The theory would be false, but not testable. A warning sign of such a situation might be the multiplication beyond plausibility of potential energy fields (of the sort discussed above) required to handle multi-component interaction. More likely, the very complexity of those T-systems in which radical emergence might be found would rule out any test of emergence. The systems of interest would just be too far from the T-constituents for any calculation based solely upon fundamental T-laws of how they should behave to be feasible. That is, of course, the situation we are in and shall remain in.

The issue of testability could become more contentious if it should turn out that the mathematical details of our best fundamental theory rule out not only analytic solutions of critical equations (a situation we are already in) but also *simulatability*[19]. It is worth remembering that the totality of physics is *not* practically testable for the simple reason that the instruments used in physical experimentation are themselves highly complex physical entities for which the hypothesis of radical emergence would have to be ruled out. The discovery and verification of properties of the most basic physical entities are the very ones that require the most complex instruments, such as particle accelerator complexes, as well as the extremely long historical chains of experimental inference which necessarily involve myriads of highly complex instruments. If it should turn out that certain complex and actual physical systems required for the testing of basic theory are *in principle* unpredictable because of certain mathematical limitations then it may be that the totality of physics is simply not a testable hypothesis at all.

The contrast between benign and radical emergence can be expressed in a theological metaphor. Imagine God creating a world. He decides it shall be made of, say, quarks and leptons that, in themselves, obey certain laws. But He has a choice about whether His new world shall be

[19] Simulatability is the feature of a theory that it is possible to calculate, in principle, the state transitions of any system in terms of the fundamental description of an initial state. Simulatability does not require that this calculation be mathematically exact; approximations are allowable so long as we can mathematically guarantee that the error of the approximation can be made as small as we like. For example, while the equations governing an isolated pendulum can be simulated by a mathematically exact representation of the system, the problem of simulating even a three-body gravitationally bound system is mathematically unsolvable. But the many-body problem can be approximated to whatever degree of accuracy we like (given arbitrarily large computing resources). There may be systems which cannot even be approximated in this sense however.

total (relative to these elementary constituents) or not. That is, He must decide whether or not to impose serious laws of emergence ‘on top of’ the properties of the basic entities. Either way, a world appears, but the worlds are different. Which world are we in? It is impossible to tell by casual inspection and perhaps impossible to tell by any experiment, no matter how idealized. Thus it may be that radical emergentism cannot be ruled out by any empirical test whatsoever, and thus it may be that we live in a world of radical emergence.

William Seager
University of Toronto at Scarborough

Appendix (Figures)

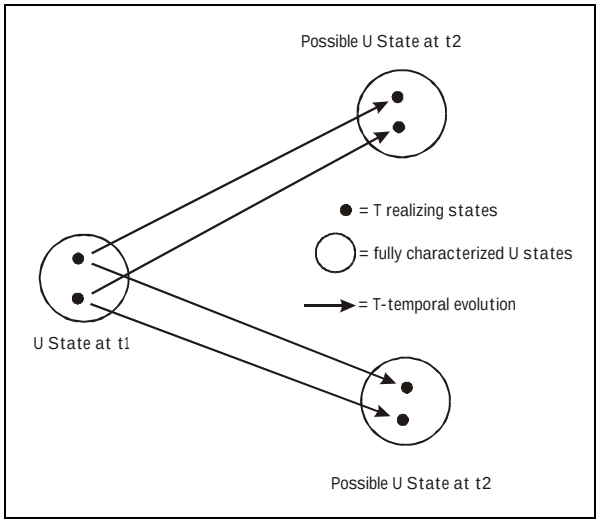

Figure 1

Go Back

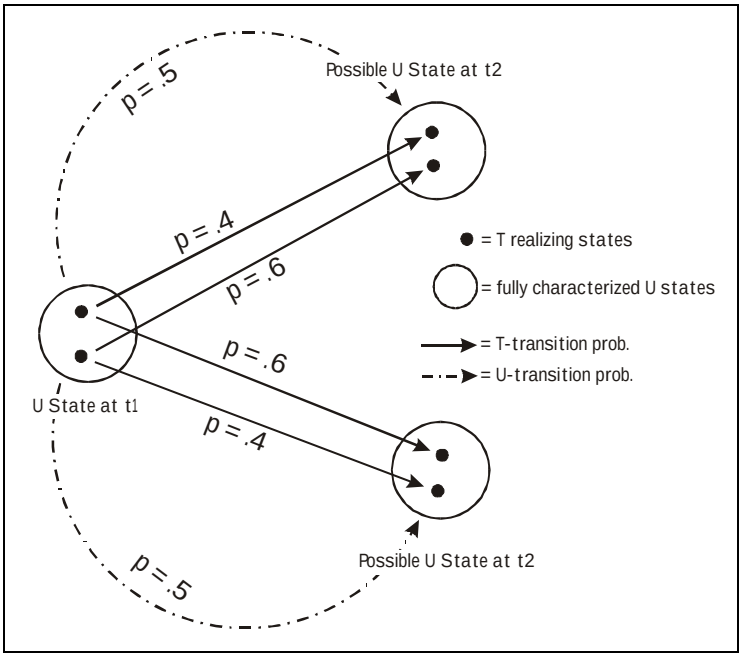

Figure 2

Go Back

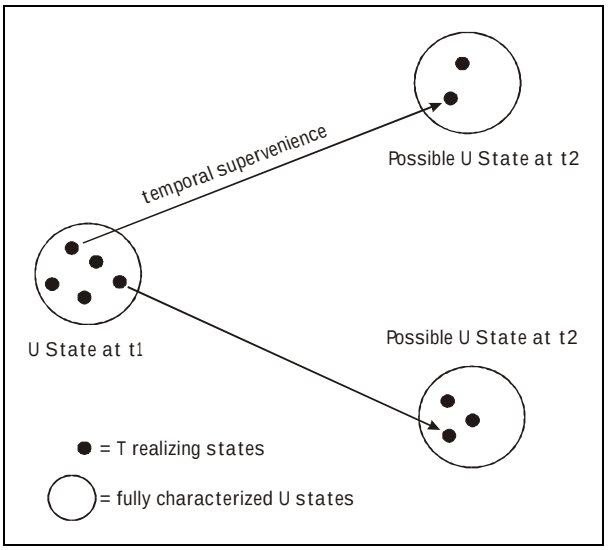

Figure 3

Go Back

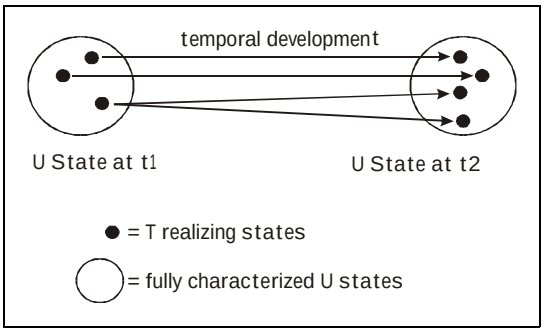

Figure 4

Go Back

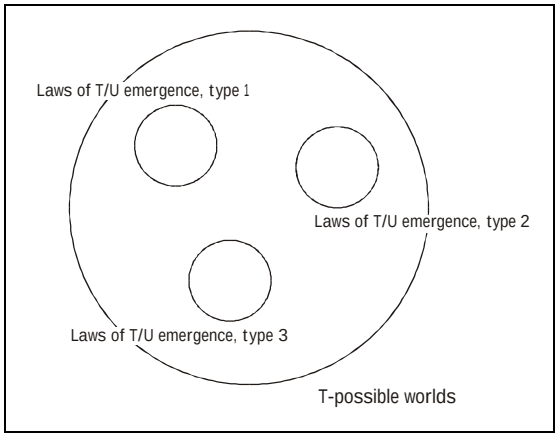

Figure 5

Go Back