ential equations "by combining element laws and continuity and compatibility equations in order to eliminate all variables except input and output" (Kutz 1998, p. 808). This is all in line with R&S's view that the behavior of a physical system can be described by dynamic systems theory and that causation is not necessary for that description (at least for some physical theories). The problem with such a description, however, is that the behavior of the entire system (for all initial conditions) is fixed by a description *without inner states* (i.e., non-input/output state variables). Yet, without inner states it is not clear how to warrant "causation talk" other than of the behaviorist kind – that would have thrown the baby out with the bath water. Worse yet, there are infinitely many different sets of equations of finitely many different non-input/output variables that give rise to the same I/O form. Put differently, there are dynamical systems that describe the behavior of a given system perfectly without having a single non-input/output variable correspond to any "natural candidate" (e.g., energy sources or energy sinks) of an "inner state" of the system (for details, see Scheutz 1999a).

The upshot of all of this is that dynamical systems theory *per se* does not, as R&S seem to suggest, provide a straightforward answer to the question whether a given physical system *realizes* a given functional architecture (Scheutz 2001), nor to the question what functional architecture(s) the system realizes (Scheutz 1999b). Kripke, for example, expresses this worry for program descriptions (i.e., that a physical machine can only "approximately" or "imperfectly" realize an infinite function) because "indefinitely many programs extend the actual behavior of the machine" (Kripke 1981, pp. 33–35). The above shows that the same is true for dynamical systems, and *a fortiori* applies to functional explanations built on or derived from them.

Note that this problem with dynamical systems is different from what R&S seem to mean by "multiple supervenience," which they take to be responsible for being able to grant supervenient properties their explanatory relevance: although the infinitely many functional architectures "induced" by the "inner state variables" all realize essentially the "same architecture" (where "same" has to be spelled out in terms of an extension of the notion of "bisimulation" defined for whole trajectories in state space instead of mere state transitions; see also Scheutz 2001), multiple supervenience seems to allow for non-bisimilar functional architectures to supervene on the same physical system – now *that* is spooky.

NOTES

1. It is also not clear what work the qualifier "broad" is supposed to do: it seems perfectly plausible that one could know all facts about feedback-driven servosystems (e.g., in the sense worked out by "control theory") and still not understand at all how these facts pertain to minds (i.e., how control states are related to mental states).

Functionalism, emergence, and collective coordinates: A statistical physics perspective on "What to say to a skeptical metaphysician"

Cosma Rohilla Shalizi

Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109. **cshalizi@umich.edu** **http://bactra.org/**

Abstract: The positions Ross & Spurrett (R&S) take on issues of information, causality, functionalism, and emergence are actually implicit in the theory and practice of statistical physics, specifically in the way it relates macroscopic collective coordinates to microscopic physics. The reasons for taking macroscopic physical variables like temperature or magnetization to be real apply equally to mental properties like pain.

Foundational questions of the kind Ross & Spurrett (R&S) worry over often don't matter much to scientists, but sometimes they matter a great deal and shape the kind of research we undertake. The answers to R&S's questions about information, causation, functionalism, and emergence matter a great deal to cognitive science. The position they argue against would inhibit not only cognitive science but also my own field of statistical physics. Moreover, we statistical physicists implicitly rely on what is essentially R&S's combination of Salmon (1984) and Dennett (1997). Therefore, I think their answers are basically right, and the skeptical metaphysician is wrong.

Consider a macroscopic physical system. It consists of many particles, each with three degrees of freedom in position, plus three in momentum, and possibly some internal degrees of freedom. (For simplicity, I'll ignore quantum mechanics.) The total number of degrees of freedom is N, and the dynamics of the system are described by equations of motion in this N-dimensional state space.

Thus far, each coordinate belongs to a particular particle. However, we are free to change our coordinate system as long as the transformation is invertible. Each degree of freedom in the new coordinates need not, and generally will not, belong to a single particle. Rather, it can be a *collective* coordinate, a function of the state of many particles, or even (like the center of mass) of all the particles (Forster 1975). The macroscopic variables that appear in physical theories are collective degrees of freedom: temperature, pressure, molecular concentrations, fluid velocity, stress, vorticity, current, and order parameters. To specify the value of one of them is to say that the system is in some particular region of the microscopic state space.

The advantage of such collective coordinates (beyond ease of measurement) is that often a fairly small number of them (m, say) interact with each other so strongly that their dynamics can be described by a deterministic evolution plus a comparatively small noise term. The noise is the effect of the remaining $N - m$ degrees of freedom on the macroscopic variables and often vanishes in the limit of large N. The macroscopic variables are then said to give a "coarse-grained" description of the system. Properly constructed, the coarse-grained variables satisfy Salmon's (1984) criteria for being a "statistical relevance basis" (Shalizi & Moore 2003). They can definitely store and transmit information over time. Moreover, they satisfy the counterfactual criteria for causality proposed by statistics and AI (Pearl 2000). The macroscopic, coarse-grained description is less precise than the microscopic one, but simpler and accurate to within a level specified by the noise. Theories in statistical mechanics start with a model of the interactions among the microscopic degrees of freedom in some system and then calculate its behavior at the coarse-grained level, including the perturbations caused by the ignored degrees of freedom (Chaikin & Lubensky 1995; Forster 1975; Keizer 1987).

Coarse-grainings that allow us to trade off complexity for accuracy are not unique. There are generally multiple levels of more or less detailed descriptions, *all* simultaneously valid for the same physical system. For instance, one can describe a fluid at a "thermodynamic" level, using quantities defined over the whole fluid, and a "hydrodynamic" one, using local currents and densities of those quantities (Keizer 1987). The thermodynamic description is a coarse-graining of the hydrodynamic one, which in turn is a coarse-graining of a more detailed molecular level. Here, one can show that the coarser levels are more predictively efficient (i.e., each bit of macroscopic information delivers more predictive information at the higher levels than the lower ones; Shalizi & Moore 2003). This gives a natural, non-mysterious definition of emergence, and one imagines it would apply nicely to mental phenomena, with (perhaps) an intentional-system level emerging from a symbolic-cognitive level, in turn emerging from a neuronal, connectionist one, and so forth down through the calcium channels to crawling molecular chaos. At each stage, we have collective coordinates of a physical system, capable of storing and transmitting information, subject to noise.

If multiple instantiation is a worry, then most of what we ordinarily consider physical quantities are in trouble. Take electric current and temperature. A current of 1 ampere can be instantiated by a certain number of electrons per second going one way, just as many hydrogen ions going the other way, and half as many calcium ions going the same way as the hydrogen, even moving "holes, propagating absences of electrons. Similarly, the property "temperature T = 300 kelvins" is instantiated by many different microphysical configurations and properties, involving momenta, spins, charges, hydrogen bonds, gravitational potentials, and so on. Many important macroscopic variables can equally well be defined as coarse-grainings or through *functional* properties relating to other macroscopic variables. An active area of statistical physics exploits the functional definitions of thermodynamic variables, abstracting ordinary thermodynamics into a purely formal structure (Ruelle 1978), and then constructing quantities that satisfy its axioms in various dynamical systems. This "thermodynamic formalism" has proved its worth in understanding chaotic dynamical systems (Beck & Schlögl 1993), hierarchical structures (Badii & Politi 1997), and turbulent flows (Chorin 1994).

To summarize, everybody agrees that things like temperature and current are physical quantities, but that they are multiply-instantiated, coarse-grained macroscopic constructions. The arguments that say mental properties are at most epiphenomenal thus apply to them, too. Against this, specifying the values of such quantities has considerable predictive power, and one can give self-contained accounts of their dynamics, subject to a certain level of noise. The extra noise and imprecision of the collective coordinates over the microscopic ones is more than offset by the gain in simplicity. They are "real patterns" (Dennett 1997). However, all this is just as true of mental properties, which are also (presumably) emergent, coarse-grained collective degrees of freedom of physical systems. There is just as much reason to treat *pain* as real and causal as to consider *electric current* so. It is not just the special sciences that need functionalism; physics needs it, too, and uses it, although we generally call it reductionism.

Protecting cognitive science from quantum theory

David Wallace

Philosophy Department, Oxford University, Oxford OX1 4JJ, United Kingdom.
david.wallace@magd.ox.ac.uk

Abstract: The relation between micro-objects and macro-objects advocated by Kim is even more problematic than Ross & Spurrett (R&S) argue, for reasons rooted in physics. R&S's own ontological proposals are much more satisfactory from a physicist's viewpoint but may still be problematic. A satisfactory theory of macroscopic ontology must be as independent as possible of the details of microscopic physics.

I find myself in close agreement with Ross & Spurrett (R&S) in the main claims of their paper; I shall confine my comments to some observations about the role which physics plays in their discussion.

R&S rightly criticise Kim's mereological definition of macroproperty for a general term like "water," but the criticism can be sharpened: Even a particular object like a table cannot really be regarded as a simple composite of non-overlapping microscopic parts. It's a tempting idea, to be sure: An extended body is just the mereological sum of its top and bottom halves; therefore, why not subdivide indefinitely until we get to the microconstituents? However, a solid object is a cloud of vastly many overlapping electron and nucleon wave functions: it is not clear even what is *meant* by saying which electron is in which spatial subregion of the object. There are ways around this problem, but they rely on dangerously strong assumptions about the present or future state of physics. (There are interpretations of quantum mechanics, for example, Bohm [1960], in which particles are something like the tiny billiard balls that philosophers treat them as – but do we really want to rest our ontology on contentious claims in quantum mechanics?)

Furthermore, even the paradigmatically "physical" properties of the object are defined not in terms of the microconstituents, but dispositionally – even the mass (!) of a solid object cannot really be defined as the sum of the masses of its atomic constituents. That algorithm gets the answer nearly right in most cases, but a helium nucleus weighs approximately 1% less than its constituents (that's why fusion works); a neutron star weighs approximately 10% less (Arnett 1996) than its constituents (that's why supernovas work). Our actual definition of mass is dispositional: Something has mass *m* if it behaves thus-and-so on the scales, or creates such-and-such a gravitational field. It is not definitional that mass is additive; it is a physical law, and only an approximate one at that.

This raises the stakes a bit, I think. R&S argue that Kim's account cannot correctly handle the natural kinds of the special sciences. However, it is actually worse: the account (I am claiming) correctly handles *hardly any macroproperty at all.*

This makes the pattern-based view of ontology espoused by Dennett (1991b), and defended by R&S, very attractive. Of course, there must be some sense in which macroscopic objects are built out of microscopic constituents and in which they are supervenient on the properties of the constituents. Dennett, by regarding macro-objects as *patterns in* the micro-ontology, rather than as *mereological sums of* that micro-ontology, provides the sort of account of compositionality that is not hostage to contentious or downright false pictures of physics.

But of course, if such an account is adopted for the whole of macro-ontology, then mental states are real in the same way that tables are real, and the causal power of the mental stands and falls with the causal power of almost everything. This would be close to a *reductio* of Kim's argument: If we are sure of anything about causation, we are sure that macroscopic objects causally influence other macroscopic objects. *Maybe* there is some esoteric notion of "causation" that applies to the ultimate microconstituents of nature only, but that notion can have little to do with "mental causation" as ordinarily understood.

Having supported R&S thus far, I wish to make one cautionary remark about their project. At times, R&S write as though the goal of a pattern ontology is to find, once and for all, the correct notion of substrate; and then define real patterns as patterns in that substrate. (This seems to be the context for their approving citation of Nottale's "fractal space-time" work; target article, sect. 4.4, para. 7) This I find dangerous: It bets our metaphysical structure on the current state of fundamental physics, despite the fact that fundamental physics frequently changes. Are "real patterns" patterns in particle distributions? Then we implicitly bet against an underlying field ontology in which particles themselves are patterns. Are "real patterns" patterns in the distribution of properties over space-time? Then we implicitly bet that space-time is fundamental (*contra* many proposals in quantum gravity) and that its role in fundamental physics is roughly the same as its role in classical physics (*contra* at least some interpretations of quantum mechanics, such as the many-worlds theory; see Wallace 2003). The danger is only heightened if we try to base metaphysics on speculative physics such as Nottale's.

One way around this problem may be to look for a sufficiently abstract characterisation of pattern as to be immune to revisions in microphysics. R&S's proposed information-theoretic approach may well succeed here, although I worry about its appeal to thermodynamic concepts like entropy: thermodynamics itself is an emergent phenomenon; therefore, there is some danger of circularity here. Another, more modest proposal would be to adopt a hierarchical view of pattern ontology: if we accept some stuff into our ontology, we should also accept patterns in that stuff. If the stuff itself turns out to be patterns in substuff, so be it. Thus, particles are patterns in the quantum field; humans are patterns in the particles; stock market crashes are patterns in the people; and so on. Such a metaphysics would be robust against, and relatively