

The physics of downward causation

Paul Davies
Australian Centre for Astrobiology, Macquarie University
New South Wales, Australia 2109

Reduction as nothing-buttery

By tradition, physics is a strongly reductionist science. Treating physical systems as made up of components, and studying those components in detail, had produced huge strides in understanding. The jewel in the crown of reductionist science is subatomic particle physics, with its recent extension into superstring theory and M theory (see, for example, Greene, 1998). The ultimate goal of these disciplines is to identify the fundamental building blocks of matter – the irreducible entities from which the entire universe is constructed.

Few would deny the efficacy of the reductionist method of investigation. The behaviour of gases, for example, would lack a satisfactory explanation without taking into account their underlying molecular basis. If no reference were made to atoms, chemistry would amount to little more than a complicated set of ad hoc rules, while radioactivity would remain a complete mystery.

As physicists have probed ever deeper into the microscopic realm of matter so, to use Steven Weinberg's evocative phrase (Weinberg, 1992), 'the arrows of explanation point downward.' That is, we frequently account for a phenomenon by appealing to the properties of the next level down. In this way the behaviour of gases are explained by molecules, the properties of molecules are explained by atoms, which in turn are explained by nuclei and electrons. This downward path extends, it is supposed, as far as the bottom-level entities, be they strings or some other exotica.

Whilst the foregoing is not contentious, differences arise concerning whether the reductionist account of nature is merely a fruitful methodology, or *whether it is the whole story*. Many physicists are self-confessed out-and-out reductionists. They believe that once the final building blocks of matter and the rules that govern them have been identified, then all of nature will, in effect, have been explained. Obviously such a *final theory* would not in practice provide a very useful account of much that we observe in the world. A final reductionist theory would not, for instance, explain the origin of life, or have much to say about the nature of consciousness. But the committed reductionist believes such inadequacies are mere technicalities, and that the *fundamental core* of explanation is captured – completely - by the reductionist theory.

A minority of physicists challenges this account of nature. Whilst conceding the power of reduction as a methodology, they nevertheless refute that the putative final theory would yield a complete explanation of the world. The anti-reductionist denies that, for example, a living cell is *nothing but* a collection of atoms, or a human being is *nothing but* a collection of cells. This, they say, is to commit the fallacy of 'nothing-buttery.' Physicists who espouse anti-reductionism usually work in fields like condensed matter physics,

where reduction often fails even as a methodology. These workers are impressed by the powerful organizational abilities of complex multi-component systems acting collectively, which sometimes lead to novel and surprising forms of behaviour.

All physicists concede that at each level of complexity new physical qualities, and laws that govern them, *emerge*. These qualities and laws are either absent at the level below, or are simply meaningless at that level. Thus the concept of wetness makes sense for a droplet of water, but not for a single molecule of H₂O. The entrainment of a collection of harmonic oscillators such as in an electrical network makes no sense for a single oscillator. The Pauli exclusion principle severely restricts the behaviour of a collection of electrons, but not of a single electron. Ohm's law finds no application to just one atom. Such examples are legion. The question we much confront, however, is *so what?* What, exactly, is it that the anti-reductionist is claiming *emerges* at each level of complexity?

In some cases the novel behaviour of a complex system may be traced in part to the fact that it is an open system. The claim that a reductive account is complete applies only to closed systems. For example, the motion of the planets in the solar system is described very precisely by Newton's laws, and even better by Einstein's theory of relativity, because the solar system is effectively isolated. Contrast this situation with, say, the motion of a hurricane, or of the great red spot of Jupiter. In this case the swirling fluids are continually exchanging matter and energy from their environment, and this leads to novel and unexpected behaviour; indeed, it may lead to random or unpredictable behaviour (the skittishness of hurricanes is notorious). It would not be possible to give an accurate account of these systems by restricting the analysis to the fluid components alone. However, there is no implication that the vortex has been seized by new forces of influences not already present in both the vortex itself and the wider environment. In principle, a satisfactory reductive account could be given by appealing to the components of the total system. But because nature abounds with chaotic systems, such a project may soon become impracticable, because it is a characteristic feature of deterministic chaos that the 'domain of influence' rapidly balloons out to encompass a vast region – even the entire universe.

Another example where limited reduction fails is quantum mechanics. A quantum superposition is famously fragile, and will tend to be rapidly degraded – decohered to use the jargon – by interactions with the environment. This is the project pursued by Zurek and reported in this volume. To illustrate this point, consider the simple case of an electron that scatters from a target with 50 per cent chance of rebounding left or right. This process may be described by a wave packet that, on its encounter with the target, splits into two blobs, one left-moving, the other right-moving. In general, there will be some overlap between the two blobs, and because the wave packet spreads as it goes, this overlap will remain as the system evolves. In practice, the electron is not isolated from its environment, and the effect of its interactions with a vast number of surrounding particles is to scramble the phases of the wave function, which results in the overlap of the blobs being driven rapidly to zero. In effect, the wave packet 'collapses' into two disconnected blobs, each representing one of the two possible outcomes of the experiment (left-moving electron and right-moving electron respectively). In this manner, the ghostly

superposition of quantum mechanics gets replaced by the classical notion of distinct states present with a 50-50 probability. In the early days of quantum mechanics, this 'collapse of the wave packet' was considered a mysterious extra process that was not captured by the rules of quantum mechanics applied to the electron and target alone. The system's decoherence and classicalization is an emergent property, and it was thought by some that this required novel new rules that were not part of quantum mechanics, but would come in at a higher level. What level? Some thought it was when the system was massive enough, others sought the rule-change at a higher level of complexity (e.g. if there was a device elaborate enough to perform a measurement of the electron's position). Some even suggested that a full understanding of the 'collapse' demanded appeal to the mind of the observer. But in fact, as Zurek and others have amply demonstrated, decoherence and wave packet collapse are well explained by appealing to the quantum interactions of the wider environment, suitably averaged over. So in this aspect of quantum mechanics at least, there is no longer any need invoke mysterious extra ingredients, or rules that emerge at the 'measurement level,' even though the 'collapse of the wave packet' is legitimately an emergent phenomenon.

What the two foregoing examples illustrate is that emergent behaviour need not imply emergent forces or laws, merely a clear understanding of the distinction between open and closed systems. And we see that language about 'the vortex' or 'the right-moving electron' is indeed merely a convenient *façon de parler* and not a reason to invoke fundamentally new forms of interaction or laws of physics. Both these examples, whilst affirming the meaningfulness of emergence as a phenomenon, nevertheless illustrate that a reductive account of that phenomenon is still adequate, so long as the environment is included within the system.

So we are confronted with the key question: is it *ever* the case that an emergent phenomenon cannot be given a satisfactory reductive account, even in principle? And if the answer is yes, then we come to the next key question: in what way, precisely, does the value-added emergent 'law' or 'behaviour' affect the system? A survey of the literature shows lots of flabby, vague, qualitative statements about higher-level descriptions and influences springing into play at thresholds of complexity, without one ever being told specifically how these emergent laws affect the individual particle 'on the ground' - the humble foot soldier of physics - in a manner that involves a fundamentally new force or law. Thus we are told that in the Bénard instability, where fluids spontaneously form convection cells, the molecules organize themselves into an elaborate and orderly pattern of flow, which may extend over macroscopic dimensions, even though individual molecules merely push and pull on their near neighbours (see, for example, Coveney & Highfield, 1995). This carries the hint that there is a sort global choreographer, an emergent demon, marshalling the molecules into a coherent, cooperative dance, the better to fulfil the global project of convective flow. Naturally that is absurd. The onset of convection certainly represents novel emergent behaviour, but the normal inter-molecular forces are not in competition with, or over-ridden by, novel global forces. The global system 'harnesses' the local forces, but at no stage is there a need for *an extra type of force* to act on an individual molecule to make it comply with a 'convective master plan.'

The fact that we need to make reference to the global circumstances to give a satisfactory account of the local circumstances is an important feature of many physical systems. It is instructive to re-cast this feature in the language of causation. We can ask, what *caused* a given water molecule to follow such-and-such a path within a given convection cell? The short answer is: the inter-molecular forces from near neighbours. But we must appeal to the *global* pattern of flow to provide a complete answer, because those near neighbours are also caught up in the overall convection. However – and this is the central point – we do not need to discuss *two sorts of forces* – near-neighbour and global forces – even though we do need to invoke two aspects in the causation story. The molecule’s motion is caused by the push and pull of neighbours, *in the context of* their own global, systematic motion. Thus a full account of causation demands appeal to (i) local forces and, (ii) *contextual information* about the global circumstances. Typically the latter will enter the solution of the problem in the form of constraints or boundary conditions.

Some emergent phenomena are so striking that it is tempting to explain them by encapsulating (ii) as a separate causal category. The term ‘downward causation’ has been used in this context (Campbell 1974). The question then arises as to whether this is just another descriptive convenience, or whether downward causation ever involves new sorts forces or influences (as was certainly the case in most versions of biological vitalism). In the cases cited above, the answer is surely no, but what about more dramatic examples, such as the mind-body interaction? Could we ever explain in all cases how brain cells fire without taking into account the mental state of the subject? If minds make a difference in the physical world (as they surely do), then does this demand additional, genuinely new, causes (forces?) operating at the neuronal level, or will all such ‘mental causation’ eventually be explained, as in the case of vortex motion, in terms of the openness of the brain to its environment and the action of coherent boundary conditions (i.e. (ii) above)?

For the physicist, the only causes that matter are, to paraphrase Thomas Jefferson, the ones that kick. Wishy-washy talk of global cooperation is no substitute for observing a real, honest-to-goodness, force that moves matter at a specific place. And if the movement is due to just the good-old forces we already know about, simply re-packaged for convenience of discussion, the response is likely to be a monumental ‘so what?’ For emergence to become more than just a way of organizing the subject matter of physics, there has to be a clear-cut example of a new type of force, or at any rate a new causative relation, and not just the same old forces at work in novel ways. Unless, that is, those forces are being subordinated in turn to some other, new, forces.

When put this bluntly, I doubt if many physicists would hold their hands on their hearts and say they believed that any such forces exist. The history of science is littered with failed forces or causative agencies (the aether, the *élan vital*, psi forces...) that try to explain some form of emergent behaviour on the cheap. In what follows I shall try to sharpen the idea of downward causation and ask just what it would take for a hard-headed physicist to be convinced that emergence demands any new causes, forces or principles beyond the routine (though possibly technically difficult) consideration of the global situation.

What is downward causation?

For the physicist, the concept of causality carries certain specific implications. Chief among these is *locality*. All existing theories of causation involve forces acting at a point in space. At a fundamental level, theories of force are expressed in terms of *local fields*, which is to say that the force acting on a particle at a point is determined by the nature of the field at that point. For example, an electron may accelerate as a result of an electric field, and the magnitude of the force causing the acceleration is given by the intensity of the electric field at the point in space the electron occupies at that moment.

When discussing the interaction between spatially separated particles, we have the concept of action at a distance, which has a non-local ring to it. The sun exerts a gravitational pull on the Earth across 150 million kilometres of space. The phenomenon may be recast in terms of local forces, however, by positing the existence of a gravitational field created by the sun. It is the action of this field on the Earth, at the point in space that the Earth happens to occupy, which creates the force that accelerates the Earth along its curved path. There is a long history of attempts to eliminate the field concept and replace it with direct non-local inter-particle action (e.g. the Wheeler-Feynman theory of electrodynamics (Davies 1995)), but these theories run into problems with physical effects propagating backward in time, and other oddities. Overwhelmingly, physicists prefer local field theories of causation.

This fundamental locality is softened somewhat when quantum mechanics is taken into account. For example, two electrons may interact and move a large distance apart. Theory suggests, and experiment confirms, that subtle correlations exist in their behaviour (see, for example, Brown & Davies, 1986; Aczel, 2002). However, it has been determined to most physicists' satisfaction that the existence of such non-local correlations does not imply a causative link between the separated particles. (A lot of popular articles convey the misconception that separated quantum particles in an entangled state can communicate information. These claims stem from confusion between correlation and communication.)

The problem of downward causation from the physicist's point of view is: How can wholes act causatively on parts if all interactions are local? Indeed, from the viewpoint of a local theory, what is a 'whole' anyway other than the sum of the parts?

Let me distinguish between two types of downward causation. The first is whole-part causation, in which the behaviour of a part can be understood only by reference to the whole. The second I call level-entanglement (no connection intended with quantum entanglement, a very different phenomenon), and has to do with higher conceptual levels having causal efficacy over lower conceptual levels.

Whole-part causation

Sometimes physicists use the language of whole-part causation for ease of description. For example, a ball rolling down a hill implies that each of the ball's atoms is accelerated according to the state of the ball as a whole. But it would be an abuse of language to say that the rotating ball *caused* a specific atom to move the way it did; after all, the ball *is* the sum of its atoms. What makes the concept 'ball' meaningful in this case is the existence of (non-local) constraints that lock the many degrees of freedom together, so that the atoms of the ball move as a coherent whole and not independently. But the forces that implement these constraints are themselves local fields, so in this case whole-part causation is effectively trivial in nature. Similar remarks apply to other examples where 'wholes' enjoy well-defined quasi-autonomy, such as whirlpools and electric circuits.

The situation is different again in the case of spontaneous self-organization, such as the Bénard instability, or the laser, where atomic oscillators are dragooned into lockstep with a coherent beam of light. But even here, the essential phenomenon can be accounted for entirely in terms of local interactions plus non-local constraints.

There are a few examples of clear-cut attempts at explicit whole-part causation theories in physics. One of these is Mach's principle, according to which the force of inertia, experienced locally by a particle, derives from the particle's gravitational interaction with all the matter in the universe. There is currently no very satisfactory formulation of Mach's principle within accepted physical theory, although the attempt to construct one is by no means considered worthless, and once occupied the attention of Einstein himself. Another example that is a bit of a grey area is the second law of thermodynamics, which states that the total entropy of a closed system cannot go down. However, 'closed system' here is a global concept. There are situations where the entropy goes down in one place (e.g. inside the refrigerator), only to go up somewhere else (the kitchen). As far as I know this law would not forbid entropy going down on Earth and up on Mars at the same instant, though one needs a relativistic theory of thermodynamics to discuss this. In quantum field theory there can be regions of negative energy that could cause a local entropy decrease, with the positive energy flowing away to another region to raise the entropy. I am not suggesting that there is an additional whole-part causation to make the respective regions 'behave themselves,' only that implicit in the second law is some sort of global constraint (or compulsion) on what happens locally.

Other examples where global restrictions affect local physics are cosmic censorship (an event horizon preventing a naked singularity) and closed timelike lines (time travel into the past) constrained by causal self-consistency. A less exotic example is Pauli's exclusion principle, where the laws governing two or more electrons together are completely different from the laws governing a single electron. It is an interesting question of whether a sufficiently long list of global restrictions would so constrain local physics as to completely define a local theory. Thus a final unifying theory of physics might be specifiable in terms of one or more global principles. However, it is important to remember that global principles do not have causal efficacy over local physics; rather,

local physics operates in such a manner as to comply with global principles. For example, we would not say that the law of conservation of energy causes a dropped ball to accelerate to the ground. The ball accelerates because gravity acts in it, but in such a way that the total kinetic and potential energy is conserved. It seems reasonable to suppose that in a final theory, all whole-part causation will reduce to local physics that happens to comply with certain over-arching global principles. In a sense, global principles may be said to *emerge* from local physics, but most physicists see things the other way round, preferring to regard global principles as somehow more fundamental.

Level entanglement

Let me now turn to the other sense of downward causation: the relationship between different conceptual levels describing the same physical system. In common discourse we often refer to higher levels exercising causal efficacy over lower. For example, mind-brain interaction: 'I felt like moving my arm, so I did.' Here the mental realm of feelings and volitions is expressed as exercising causal efficacy over flesh. Another example is hardware versus software in computing. Consider the statement: 'The program is designed to find the smallest prime number greater than one trillion and print out the answer.' In this case the higher-level concept 'program' appears to call the shots over what an electronic gizmo and printer and paper does. Many examples may be found in the realm of human affairs, such as economics. Pronouncements such as, 'stock market volatility made investors nervous' conveys the impression that the higher-level entity 'the stock market' in part determines how individual agents behave.

In the latter two examples at least no physicist would claim that there are any mysterious new physical forces acting 'down' from the software onto the electronic circuitry, or from the stock market onto investors. Software talk and reference to 'market forces' in economics do not imply the deployment of additional *physical* forces at the component level. The existing inventory of physical forces suffices to account for the detailed behaviour of the components. Once again, the best way to think about downward causation in these examples is that the global system harnesses existing local forces. The mind-brain example is much harder because of the complexity and openness of the system. A more dramatic example of mind-brain causation comes from the field of neurophysiology. Recent work by Max Bennett (Bennett & Barden, 2001) in Australia has determined that neurones continually put out little tendrils that can link up with others and effectively rewire the brain on a time scale of 20 minutes! This seems to serve the function of adapting the neuro-circuitry to operate more effectively in the light of various mental experiences (e.g. learning to play a video game). To the physicist this looks deeply puzzling. How can a higher-level phenomenon like 'experience,' which is also a global concept, have causal control over microscopic regions at the sub-neuronal level? The tendrils will be pushed and pulled by local forces (presumably good old electromagnetic ones). So how does a force at a point in space (the end of a tendril) 'know about,' say, the thrill of a game?

Twenty years ago I conceived of a device to illustrate downward causation in a straightforward way (Davies, 1986). Consider a computer that controls a microprocessor

connected to a robot arm. The arm is free to move in any direction according to the program in the computer. Now imagine a program that instructs the arm to reach inside the computer's own circuitry and rearrange it, e.g. by throwing a switch or removing a circuit board. This is software-hardware feedback, where software brings about a change in the very hardware that supports it. In a less crude and brutal formulation of this scenario we might imagine the evolution of the computer/arm to be quite complex, as the continually rearranged circuitry changed the instructions to the arm, which in turn changed the circuitry...

Although it is hard to think of this example in terms other than software acting on hardware, there presumably exists a complete hardware description of events in terms of local interactions. In other words, there are no new forces or principles involved here. Use of terms like software and arm are simply linguistic and conceptual conveniences and not causal categories.

Which way do the arrows of causation point?

An interesting example of downward causation is natural selection in evolution. Here the fate of an organism, maybe an entire species, depends on the circumstances in the wider ecology. To take a specific example, consider the case of convergent evolution, where similar ecological niches become filled by similar organisms, even though genetically these organisms might be very far apart. The eye has evolved in at least 40 independent ways in insects, birds, fish, mammals etc.; although the starting points were very different, the end products fulfil very similar functions. Now the morphology of an organism is determined by its DNA, specifically by the exact sequence of base pairs in this molecule. Thus one might be tempted to ask, how does the biosphere act downwards on molecules of DNA to bring about species convergence? But this is clearly the wrong question. There is no mystery about convergence in Darwinian evolution. Random mutations alter the base-pair sequences of DNA and natural selection acts as a sieve to remove the less fit organisms. Selection takes place at the level of organisms, but it is the genes (or base-pair sequences) that get selected.

Darwinism provides a novel form of causation inasmuch as the causal chain runs counter to the normal descriptive sequence. Chronologically, what happens is that first a mutation is caused by a local physical interaction, e.g. the impact of a cosmic ray at a specific location with an atom in a DNA molecule. Later, possibly many years later, the environment 'selects' the mutant by permitting the organism to reproduce more efficiently. In terms of physics, selection involves vast numbers of local forces acting over long periods of time, the net result of which is to bring about a long-term change in the genome of the organism's lineage. It is the original atomic event in combination with the subsequent complicated events that together give a full causative account of the evolutionary story. Yet biologists would be hard-pressed to tell this story in those local physical terms. Instead, natural selection is described as having causal powers, even though it is causatively neutral – a sieve. In this respect, natural selection is better thought of as a constraint, albeit one that may change with time.

Information and level-entanglement

There is one place in mainstream physics where two conceptual levels seem to become inextricably entangled in our description of events, and that is quantum mechanics. (Recall that I am not using the word entanglement here in the conventional sense of an entangled quantum state.) The much-vaunted wave–particle duality of quantum mechanics conceals a subtlety concerning the meaning of the terms. Particle talk refers to hardware: physical stuff such as electrons. By contrast, the wave function that attaches to an electron encodes *what we know* about the system. The wave is not a wave of ‘stuff,’ it is an information wave. Since information and ‘stuff’ refer to two different conceptual levels, quantum mechanics seems to imply a duality of levels akin to mind-brain duality.

When an observation is made of a quantum system such as an electron, the wave function typically jumps discontinuously as our information about the electron changes. For example, we may measure its position which was previously uncertain. Thereafter the wave evolves differently because of the jump. This implies that the particle is likely to be found subsequently to be moving differently from the manner in which it might have been expected to move had the measurement not been made. Quantum mechanics appears to mix together information and matter in a bewildering way.

What I have been describing is really an aspect of the famous measurement problem of quantum mechanics, which is how we should understand the ‘jump’ referred to above. The work of Zurek and others (Zurek 1982) attempts to eliminate the appearance of a discontinuity and the intervention of an observer by tracing the changes to the electron’s wave function to a decohering environment. However, even if the wave function is seen to evolve smoothly, it must still be regarded as referring to knowledge or information about the quantum system, and information is meaningful only in the context of a system (e.g. a human observer) that can interpret it. Wheeler has stressed the level-entanglement involved in quantum mechanics in his famous ‘meaning circuit,’ in which ‘observership’ underpins the laws of physics (for a recent review see Barrow, Davies & Harper, 2003). We can trace a causal chain from atom through measuring apparatus to observer to a community of physicists able to interpret the result of the measurement. In Wheeler’s view there must be a ‘return portion’ of this ‘circuit’ from observers back down to atom.

Information enters into science in several distinct ways. So far, I have been discussing the wave function in quantum mechanics. Information also forms the statistical basis for the concept of entropy, and thus underpins the second law of thermodynamics (information should not come into existence in a closed system). In biology, genes are regarded as repositories of information – genetic databanks. In this case the information is semantic; it contains coded instructions for the implementation of an algorithm. So in molecular biology we have the informational level of description, full of language about constructing proteins according to a blueprint, and the hardware level in terms of molecules of specific atomic sequences and shapes. Biologists flip between these two modes of description without addressing the issue of how information controls hardware – a classic case of downward causation. There is a fourth use of information in physics,

entering via the theory of relativity. This says that information shouldn't travel faster than light.

Recently there have been ambitious attempts to ground all of physics in information; in other words, to treat the universe as a gigantic informational or computational process (Frieden, 1998). An early project of this type is Wheeler's 'It from bit' proposal (Barrow, Davies & Harper, 2003). We might call this 'level inversion' since information is normally regarded as a higher-level concept than, say, particles.

Where do we go from here?

Most physicists are sceptical of downward causation, because they believe there is 'no room' in existing theories of causation for additional forces. Certainly the idealised model of a physical system – a closed deterministic dynamical system obeying local second order differential equations – is causally closed too. Sometimes quantum mechanics, with its inherent indeterminism, is seen as opening a chink through which additional forces or organizational influences might enter. This is the reasoning behind some attempts to root consciousness and freewill in quantum fuzziness (see, for example, Penrose 1989, Hodgson, 1991). However, standard quantum mechanics is really a deterministic theory in its dynamics, even though its predictions are statistical. Slipping in extra forces by 'loading the quantum dice' is an unappealing prospect. For one reason, the 'loading forces' would by definition lie outside the scope of quantum mechanics, leading to a dualistic description of nature in which quantum and non-quantum forces acted in combination. But quantum mechanics makes no sense if it is not a universal theory. If control could be gained over the 'loading forces' they could, for example, be used to violate the uncertainty principle.

Another way to escape the strictures of causal closure is to appeal to the openness of some physical systems. As I have already stressed, top-down talk refers not to vitalistic augmentation of known forces, but rather to the system harnessing existing forces for its own ends. The problem is to understand how this harnessing happens, not at the level of individual intermolecular interactions, but overall – as a coherent project. It appears that once a system is sufficiently complex, then new top-down rules of causation *emerge*. Physicists would like to know whether these rules can ultimately be derived from the underlying laws of physics, or must augment them. Thus a living cell commandeers chemical pathways and intermolecular organization to implement its plan encoded in the genome. The cell has room for this supra-molecular coordination because it is an open system, so its dynamics is not determined from within the system. But openness to the environment merely explains *why* there may be room for top-down causation; it tells us nothing about *how* that causation works.

Let me offer a few speculations about how. In spite of the existence of level entanglement in quantum physics and elsewhere, none of the examples cited amounts to the deployment of specific local forces under the command of a global system, or subject to emergent rules at higher levels of description. However, we must be aware of the fact that physics is not a completed discipline, and top-down causation may be something that

would not show up using current methods of enquiry. There is no logical impediment to constructing a whole-part dynamics in which local forces are subject to global rules. For example, it is straightforward to design a cellular automaton in which the evolution rules for a given pixel are determined by a global variable, such as some measure of the complexity of the entire pixel array. There have even been suggestions (Davies, 1987; Leggett, 1994) to introduce complexity as a physical variable in quantum systems, so that the rules for the evolution of a complex system might depart from those of a simple system. Since complexity is another higher-level concept and another global variable, this would introduce explicit downward causation into physics. To use the now-discredited terminology of the quantum measurement problem, one might posit that the wave function ‘collapses’ when the system of interest (e.g. the electron) couples to an environment that is sufficiently complex. There have been attempts to introduce non-linearity into quantum mechanics to explain the ‘collapse,’ but as far as I know there is no mathematical model of system complexity entering the dynamics of a complex system to bring about this step. My proposal evades the problems associated by the ‘loading forces’ suggestion discussed above, because it operates at the interface between the quantum and classical realms. It is easier to imagine downward causation acting at that interface rather than mingling with quantum processes deep within the quantum realm.

Any attempt to introduce explicitly global variables into local physics would necessarily come into conflict with existing purely local theories of causation, with all sorts of ramifications. First would be consistency with experiment. If downward causation were limited to complexity as a variable, then effects would most likely be restricted to complex systems, where there is plenty of room for surprises. For example, in the case of the living cell it is doubtful if additional ‘organizational’ forces related to a global complexity variable acting at the molecular level would have been detected by techniques used so far. Similar remarks apply to mind-brain causation. Second, global principles such as the second law of thermodynamics might be affected by downward causation. For example, a cellular automaton in which the dynamical rules depend in certain ways on the complexity of the state might develop entropy-lowering behaviour, and thus be ruled out of court.

Finally let me discuss a different mechanism of downward causation that avoids the problem of coming into conflict with existing local theories. As remarked already, many authors have suggested that the universe should be regarded as a gigantic computer, or information-processing system, and that perhaps information is more primitive than matter, underpinning the laws of physics. As pointed out long ago by Landauer (Bennett & Landauer 1985), the information-processing power of the universe is limited by its resources, specifically, by the number of degrees of freedom contained within the particle horizon (the causal limit of the universe imposed by the finite speed of light). As the universe ages, so the particle horizon expands, and more and more particles come into causal contact. So the universe begins with very limited information-processing power, but its capability grows with time. Seth Lloyd (2002) has estimated the maximum amount of information that the universe has been able to process since the big bang. The answer comes out to be about 10^{120} bits. Now this number 10^{120} is very familiar. It turns out to be the same factor by which the so-called cosmological constant is smaller than its ‘natural’

value as determined on dimensional grounds. (The cosmological constant refers to a type of anti-gravitational force that seems to be accelerating the rate of expansion of the universe.) This vast mismatch between the observed and actual values of the cosmological constant was described by Stephen Hawking as ‘the biggest failure of dimensional analysis known to science.’ The mismatch is known as ‘the cosmological constant problem.’

A possible solution of the cosmological constant problem comes from top-down causation. Suppose this quantity, normally denoted Λ , is not a constant at all, but a function of the total amount of information that the universe has processed since the beginning. Lloyd points out that the processed information increases like the square of the age of the universe, t^2 . Then I hypothesise

$$\Lambda(t) = \Lambda_{\text{Planck}} (t_{\text{Planck}}/t)^2$$

where ‘Planck’ refers to the Planck time, 10^{-43} s, at which the universe contains just one bit of information. It can be seen from the above equation that Λ starts out very large, then decays with time, dropping to its present value and declining still further in the future. This, then, is a theory where a basic force of nature derives (via a mechanism that I have not attempted to explicate) from the higher-level quantity ‘processed information,’ in a manner that leads to directly observable consequences.

Reductionist local causation has been a feature of physics since Newton made the fundamental separation between dynamical states and laws. Attempts to include explicit whole-part causative processes would introduce a fundamental change in theoretical physics by entangling law and state in a novel manner. Whether this complication would be welcomed by physicists is another matter.

Many emergentists would not welcome it either. The conventional emergentist position, if one may be said to exist, is to eschew the deployment of new forces in favour of a description in which existing forces merely act in surprising and cooperative new ways when a system becomes sufficiently complex. In such a framework, downward causation remains a shadowy notion, on the fringe of physics, descriptive rather than predictive. My suggestion is to take downward causation seriously as a causal category, but it comes at the expense of introducing either explicit top-down physical forces or changing the fundamental categories of causation from that of local forces to a higher-level concept such as information.

Footnotes

Aczel, Amir (2002) *Entanglement* (Four Walls Eight Windows, New York).

Barrow, John, Davies, Paul & Harper, Charles (2003) (eds.) *Science and Ultimate Reality* (Cambridge University Press, Cambridge).

Bennett, C. & Landauer, R. (1985) ‘The fundamental physical limits of computation,’ *Scientific American*, July 1985, pp. 48-56.

- Bennett, M.R. & Barden, M.R. 'Ionotropic (P2X) receptor dynamics at single autonomic varicosities,' *NeuroReport* **12**, A91-A97 (2001).
- Brown, Julian & Davies, Paul (1986) *The Ghost in the Atom* (Cambridge University Press, Cambridge).
- Campbell, D. T. (1974) '“Downward causation” in Hierarchically Organized Biological Systems', in *Studies in the Philosophy of Biology* (eds. F.J. Ayala & T. Dobzhansky; Macmillan, London), p. 179-186.
- Coveney, Peter & Highfield, Roger (1995) *Frontiers of Complexity* (Ballantine, New York), Chapter 6.
- Davies, Paul (1986) 'Time asymmetry and quantum mechanics,' in *The Nature of Time* (eds. Raymond Flood & Michael Lockwood; Blackwell, Oxford), Chapter 7.
- Davies, Paul (1987) *The Cosmic Blueprint* (Simon & Schuster, New York).
- Davies, Paul (1995) *About Time* (Simon & Schuster), Chapter 9.
- Frieden, B. Roy (1998) *Physics from Fisher Information* (Cambridge University Press, Cambridge).
- Greene, Brian (1998) *The Elegant Universe* (Norton, New York).
- Hodgson, David (1991) *The Mind Matters* (Oxford University Press, Oxford).
- Leggett, A.J. (1994) *Problems of Physics* (Oxford University Press, Oxford).
- Lloyd, Seth (2002) 'Computational capacity of the Universe,' *Physical Review Letters*, **88**, 237901.
- Penrose, Roger (1989) *The Emperor's New Mind* (Oxford University Press, Oxford).
- Weinberg, Steven (1992) *Dreams of a Final Theory* (Pantheon, New York), Chapter 3.
- Zurek, W.H. (1982) 'Environment-induced superselection rules,' *Phys. Rev.* **D26**, 1862.