12

The Grain Problem[1]

Michael Lockwood

> Think of what consciousness feels like, what it feels like at this moment. Does that *feel* like billions of tiny atoms wiggling in place?
>
> (Carl Sagan[2])

> How can technicolour phenomenology arise from soggy grey matter?
>
> (Colin McGinn[3])

There is, today, no glimmer of a consensus amongst philosophers about the mind–body problem. Nevertheless, an increasing number of philosophers find themselves occupying a middle ground between physicalist reductionism, on the one hand, and dualism on the other. Physicalist reductionism I take to be the view that the physical story about what is going on in the brain and the world with which it interacts is in some sense the whole story. If there really are such things as mental states and processes—which eliminative materialists notoriously deny—then their existence must be logically implicit in facts statable in the language of physics. Space does not permit a detailed rebuttal of reductionist physicalism; nor do the arguments I have elsewhere presented[4] admit of brief summary. But the simple intuitive argument is that a being provided with a description of you or me couched purely in the language of physics—even if it possessed unlimited powers of ratiocination—would have no way of deducing that our bodies were associated

[1] In writing this article, I have benefited greatly from an excellent critique of my views—as set out in ch. 10 of my *Mind, Brain and the Quantum: The Compound 'I'* (Oxford, 1989)—which appears in J. A. Foster's *The Immaterial Self* (London, 1991), 119–30. My statement of the grain problem, in particular, owes much to this discussion.

[2] *Contact: A Novel* (New York, 1985), 255.

[3] 'Can We Solve the Mind–Body Problem?' *Mind*, 98 (1989), 349.

[4] *Mind, Brain and the Quantum*, ch. 8.

with awareness at all, much less what specifically it was *like* to be you or me.[5] There is, of course, a lot more to be said on the matter; but attempts to disarm such intuitive arguments seem to me, in the end, uniformly unsuccessful. Indeed, for those not blinded by science, the falsity of reductionist physicalism will probably seem almost too obvious to require argument: Galen Strawson aptly describes it as 'moonshine'.[6]

Dualism, on the other hand, is unattractive to most philosophers because embracing such a doctrine seems more like giving up on the mind–body problem than providing a genuine solution to it. Dualism does little or nothing to satisfy our cravings for an integrated world view. It remains obscure, on the dualist theory, just how the material is supposed to dovetail with immaterial mind. For, on the face of it, there are no mind-shaped gaps in the material fabric; the material world offers no explanatory or descriptive slots into which immaterial minds could comfortably fit. (One pictures matter saying to Cartesian mind: 'This universe ain't big enough for both of us'!)

Anyway, I shall be assuming in this paper that, though reductionist physicalism is false, some form of materialism is nevertheless true. Conscious states and events are, on the view I favour, states of, or events within, the brain. But the very existence of consciousness shows that there is more to the matter of the brain (and hence presumably to matter in general) than is currently capable of being captured in the language of physics or physiology. How, then, is this 'more' to be conceived? Well, Bertrand Russell suggested, in the 1920s, that, in respect of the brain, awareness might be providing content, where science provides only form.[7] All that we really know of the physical world, on the basis either of sense perception or of physical theory, Russell argued, is that it possesses a certain *causal structure*. Any attribute of a physical system, whether it be shape, size, or electric charge, is really known to us only as whatever it is that occupies a certain logical niche within a causal-explanatory system. We have no way of knowing what the external world is like *in itself*; its intrinsic

[5] See T. Nagel, 'What is it Like to be a Bat?', *Philosophical Review*, 83 (1974), 435–50; repr. in id., *Mortal Questions* (Cambridge, 1979), 165–80.

[6] G. Strawson, 'Consciousness, Free Will, and the Unimportance of Determinism', *Inquiry*, 32 (1989), 3.

[7] See esp. B. Russell, *The Analysis of Matter* (London, 1927).

character is systematically hidden from the gaze of ordinary observation or experiment. But now, the brain is itself a part of the physical world, and we are assuming that conscious states are brain states. We certainly seem to know, from introspective awareness, the intrinsic character of an itch or the sound of middle C, played on the piano, or a patch of phenomenal yellow. So if conscious states *are* brain states, do we not here have a corner of the physical world whose intrinsic nature precisely *is* made manifest to us, albeit in a very limited way? This was Russell's suggestion: that in consciousness, a fragment of physical reality is, so to speak, being apprehended from within.

This idea—which seems to me the only approach to the philosophical mind–body problem, currently on offer, that holds out the slightest promise—can be thought of as a neat inversion of a celebrated theory put forward some thirty years ago by J. J. C. Smart. Smart suggested that mental state terms were, as he put it, 'topic neutral'. According to Smart, when I say that I am experiencing a yellowish-orange patch in my visual field, I am saying something like this: 'There is something going on which is like what is going on when I have my eyes open, am awake, and there is an orange illuminated in good light in front of me, that is, when I really see an orange.'[8] This then leaves it open for the physiologist to discover what, in relevant respects, actually is going on under such conditions, physiologically speaking, and identify it with the occurrence of phenomenal yellow-orange. But of course this isn't at all what I am saying when I report that I am experiencing phenomenal yellow-orange; if it were, it would follow, absurdly, that there was nothing to prevent a congenitally blind person from having as rich and complete an understanding of such introspective reports as a sighted person. Russell's view turns this unworkable theory on its head: for him it is the *physical* descriptions, rather than the mental ones, which are topic neutral.

It is at this point that we encounter the *grain problem* (a difficulty attributed to Wilfrid Sellars[9]). For if the immediate objects of introspective awareness just are states of, or events within, the brain, seen as they are in themselves, why do they *appear to be* so

[8] J. J. C. Smart, 'Sensations and Brain Processes', *Philosophical Review*, 68 (1959), 141–56.

[9] W. Sellars, 'The Identity Approach to the Mind–Body Problem', *Review of Metaphysics*, 18 (1965), 430–51.

radically different from anything that a knowledge of the physiology of the brain would lead one to expect?

That rather vague intuitive thought may be resolved into three more specific difficulties, each of which can be regarded as an aspect of the grain problem, as I conceive it. First is the fact that the phenomenal objects of introspective awareness are far less finely structured than are any plausible physiological correlates. Consider, for example, a phenomenally flawless auditory experience, of a note, say, on a violin. Its physiological substrate, presumably, is a highly structured, not to say messy, concatenation of changes in electrical potential within billions of neurons in the auditory cortex, mediated by the migration of sodium and potassium ions across cell membranes, and of molecules of transmitter substances within the chemical soup at the synapses. How do all these microstructural discontinuities and inhomogeneities come to be *glossed over*, in such a way as to generate the elegant perfection of auditory phenomenology that we associate with the playing of a Yehudi Menuhin? How are we to make philosophical sense of such phenomenological *coarse-graining*?

The second problem is that the structure we do encounter at the phenomenal level seems not to match, even in coarse-grained fashion, that of the underlying physiology, as revealed by scientific investigation. The phenomenal contents of awareness don't appear to have the *right kind* of structure; what is ostensibly lacking, here, is even the most approximate isomorphism between states of awareness and the underlying physiological goings-on that, on my view, they are supposed to be mirroring. In particular, three-dimensional spatial arrangement, and changes therein, seem central to all physical structure. Where, then, are their phenomenological counterparts? Of course, there is the visual field, and auditory and somatic-sensory space. But these are local, modality-specific *representations*, merely, of regions of the external world. We search in vain for some global, overarching mode of phenomenological organization that could plausibly be equated with introspectively encountered spatial layout. It is all very well to insist that the scientist's characterization of the brain, as of the physical world in general, is ultimately topic neutral; so that the terms of the characterization are, in the final analysis, mere placeholders for unspecified intrinsic natures. The problem is that the phenomenal pegs, as John Foster neatly puts it, seem

not to be the right shape to fit these holes in the topic-neutral characterization.[10]

Someone may see in these difficulties an argument for functionalism. The functionalist would regard the relation between a phenomenological description of the contents of consciousness and a physiological description of the corresponding brain-processes as analogous to that between a description of the workings of a computer in software terms, on the one hand, and in terms, say, of the electronic configuration of the underlying circuits, on the other. Thus, brain states, for the functionalist, impinge on awareness only *qua* possessors of certain high-level causal-functional roles. Precisely what, in physiological terms, are playing those roles, and how they do so, is, at the level of phenomenology, essentially irrelevant.

Functionalism, however, has its own problems—most notably its inability to explain why functional roles should be associated with any phenomenal qualities—*qualia*—at all. And in any case, it would seem, intuitively, perfectly possible for there to be a system functionally equivalent to a human mind, in which the corresponding functional roles were associated with different *qualia* from those associated with these roles in our own case.[11] Functionalism may have some plausibility in accounting for mental structure but, on the face of it, fails utterly to account for phenomenal *content*. Moreover, all arguments one could mount against reductionist physicalism apply *a fortiori* to functionalism; since if functionalism were true, reductionist physicalism clearly *could be* true also. If a physical system is, so to speak, running the right programs, then it follows, for the functionalist, that it has certain mental states; and this is something that a being with sufficient ratiocinative power could presumably read off from a description of the system couched in the language of physics. If, as I have been suggesting, reductionist physicalism is essentially a non-starter, then so too is functionalism—at least if put forward as a global theory of mind.

The third aspect of the grain problem that I wish to consider is raised by the profligate *qualitative diversity* of the phenomenal

[10] Foster, *The Immaterial Self*, 126.

[11] See N. Block, 'Troubles with Functionalism', in *Minnesota Studies in the Philosophy of Science*, 9, ed. C. W. Savage, Minneapolis, 1978), 261–325, and also my *Mind, Brain and the Quantum*, ch. 3.

realm, which seems flatly at odds with the comparative qualitative homogeneity of the physical ingredients out of which any corresponding brain state could realistically be composed. There are two levels at which this might be argued. Both visual and auditory information, according to the current wisdom, are encoded—albeit in different parts of the brain—by firing rates within certain batteries of neurons. But there is (as far as I am aware) nothing qualitatively distinctive about a neuron in the auditory cortex, or the corresponding action potential, to mark it out from a neuron, or the firing of a neuron, in the visual cortex. So how, on this basis, is one to account, say, for the fundamental phenomenological difference between a sound and a flash?

The other level at which the point could be argued is that of particle physics. The most promising currently available candidate for a so-called *theory of everything* (TOE) is something known as *superstring theory*.[12] According to this theory, everything is ultimately composed of incredibly minute loops—the 'strings'—with length and tension, but no thickness; everything that happens is ultimately a matter of the motion and interaction of these strings; elementary particles are strings in different vibratory states. These strings are held to inhabit a ten-dimensional space-time, in which six of the spatial dimensions are curled up in such a tight radius that they are effectively undetectable *as spatial dimensions*, though their presence manifests itself in the form of forces. The details of the theory scarcely matter, for our purposes. What does matter is that, once again, it seems incomprehensible that different combinations of collective or individual string states could generate the qualitative diversity that is manifest at the phenomenal level. It seems inconceivable in much the same way, and for much the same reasons, that it is inconceivable that an artist, however skilled, should conjure the simulacrum of a Turner sunset from a palette containing only black and white paints.

What is ostensibly lacking, both at the neuronal level and at the level of particle physics, is, most obviously, the requisite qualitative potential—just as black and white paints provide the potential for an infinite number of shades of grey, but not for a yellow or a red. But there is also (as John Foster has pointed out[13]) a subtler

[12] See M. B. Green, 'Superstrings', *Scientific American*, 255 (Sept. 1986), 44–56.
[13] Foster, *The Immaterial Self*, 127–8.

difficulty having to do with the possibility of securing, at the fundamental level, the required qualitative *flexibility*. One might, in speculative vein, attempt some wholesale enrichment of the physical microstructure—crediting the basic ingredients of the physicist's ontology with intrinsic attributes way beyond what are called for by their explanatory roles within physical theory, but which are specifically tailored to the demands of phenomenology. The trouble then, however, is that it seems scarcely deniable that, at some level, these fundamental ontological ingredients, whatever they are, must be broadly *interchangeable*. What, one may ask, is the use of attributing, say, embryonic colour to the ultimate physical components involved in the neuronal goings-on that are supposed to be constitutive of a phenomenal patch of red, if these self-same constituents are also to be capable of figuring in auditory or olfactory experiences which are wholly devoid of visual phenomenology? Little is gained if what one does in order to account for the *presence* of phenomenal qualities in one place has the effect of making a mystery of their ostensible *absence* elsewhere.

With regard to the first of these three difficulties, a concrete analogy may help to fix ideas. Consider a (monochrome) newspaper photograph. Seen at very close quarters, or through a magnifying glass, it stands revealed as a rectangular array of different-sized black dots on a white background. But casual inspection shows, rather, lines, edges, and patches of black, white, and varying shades of grey. Let the latter appearance correspond, in our analogy, to the phenomenal aspects of an experience, and the array of dots to the nitty-gritty of ion exchange and so forth, which is constitutive of the corresponding brain-process.

The very word 'introspection' invokes a supposed analogy with perception: the metaphor of the 'inner eye'. (Compare Kant's talk of an 'inner sense', complementary to the 'outer senses'.) Now if there really were a close parallel here, this first aspect of the grain problem would scarcely be troubling. Just as, with the photograph, the limited resolving power of the eyes ensures that, if we stand back sufficiently, we shall have the illusion of continuity, so we could envisage the mind, in introspection, as standing back from the underlying brain-processes—again, with consequent loss of resolution. Particulate and discontinuous physico-chemical activity will yield perceived continuity, just as the discrete patches of ink

on paper give way to ostensibly continuous lines and patches of black, white, and grey. But of course, this picture is simply incoherent. For the mind is not supposed to exist *over and above* the relevant brain activity. And no literal sense can be attached to the notion of the conscious mind being distanced, in this fashion, *from itself*.

Coarse-graining within ordinary perception is ultimately to be explained via the concept of a *mental representation*. It is a mental representation of the external object, rather than the object itself, that is directly before the mind in ordinary perception. And this mental representation is linked to the external object by an information-conveying causal chain. Degree of resolution is largely a matter of *how much* information about the external object is conserved in transmission; though, more generally, it is also a matter of how the information is encoded and reprocessed. (Thus, 'smoothing' of the data is presumably, in part, a product of specific processing; it could hardly be accounted for on the basis merely of information degradation.)

But, as I say, there is no such story to be told in regard to introspective awareness. Introspection is not a distinct sensory modality whose objects differ from those of 'outer sense' by being internal instead of external to the conscious mind. Rather, it is distinguished by one's cognitive or intentional *focus*. Thus, any of the ordinary five senses may be exercised in introspective mode; and doing so is a matter of taking as one's cognitive focus the mental representations themselves, instead of the external objects (if any) which they represent. (Compare the way in which, while watching the Wimbledon men's finals on television, one could switch one's mental focus from the players themselves to the corresponding images on the screen—in the context, say, of wondering whether one should adjust the contrast.) Hence, there are no distinctively introspective meta-mental representations, which stand to introspection as do ordinary visual, auditory, etc. representations to sight and hearing—and whose separation from their mental objects could help us resolve this aspect of the grain problem. And even if there were, the original problem would simply re-emerge at the level of these meta-representations themselves. Our difficulties begin at the point where the perceptual buck stops.

The force of these arguments will, I suspect, be lost on some

people. Clearly, someone might protest, there are macroscopic qualities, and there is macroscopic structure: consider liquidity, for example, or temperature, or sphericity. These are perfectly genuine features of physical reality; so why shouldn't it be correspondingly macroscopic features of brain activity that manifest themselves in awareness? But macroscopic features such as those cited are not genuinely *emergent* attributes of the physical world. On the contrary, high-level descriptions like 'liquid', 'hot', or 'spherical' apply—so it would seem—entirely in virtue of what holds true at the microlevel. And if so, it appears to follow that external physical reality can, in thought and perception, present itself to the mind in such high-level terms only by courtesy of the mediating role of mental representations.

I am not, of course, suggesting that the objects of direct awareness come unconceptualized. Thus the presence, within one's visual field, of a number of black dots—even if, in contrast to the dots in our newspaper photograph, they are individually perceived as such—may inescapably carry with it the interpretation *circle*. But that does nothing to explain how what is presented to awareness can, in another instance, just *be* a phenomenally continuous circle, when the physical substrate of the experience consists of a discontinuous array of, say, discrete centres of electrical activity.

Grover Maxwell (whose statement of the grain problem is the must lucid I have come across in the published literature) suggests that, if we are looking for physical structure that is isomorphic to the phenomenal structure encountered in awareness, we might find it at what he dubs the 'middle-sized' level.[14] What he has in mind is a level of structure intermediate between, and less familiar than, quantum microstructure and quasi-classical macrostructure: a level the better understanding of which might, he thinks, hold the key to the elusive goal of bringing together, into a consistent whole, quantum mechanics and general relativity. But there is a fundamental philosophical unclarity in Maxwell's proposal. For what exactly is 'middle-sized' structure supposed to consist in? Is it supposed to be structure which is, in the above sense, *high-level* with respect to electrons and the like—albeit low-level with respect

[14] G. Maxwell, 'Rigid Designators and Mind–brain Identity', in *Minnesota Studies in the Philosophy of Science*, 9, ed. C. W. Savage (Minneapolis, 1978), 399.

to, say, blizzards, buffaloes, ball-bearings, and bacteria, hamsters, ham sandwiches, and housing estates? If so, then all he's really talking about—so it's tempting to argue—is microstructure under a (relatively) high-level description. And all the considerations invoked in the past few paragraphs still apply; it will remain a complete mystery how direct introspective contact with brain activity—unmediated by intervening mental representations—can reveal middle-sized structure to the total exclusion of the microstructure which is ultimately constitutive of it.

Perhaps, however, what Maxwell means by middle-sized structure is not merely high-level structure, with respect to the quantum microstructure, but something genuinely *emergent*, in a sense in which liquidity, temperature, and the like are not. The only sense I can attach, in the present context, to Maxwell's middle-sized structure being emergent is that it is structure which is instantiated—in part or in whole—by *emergent qualities*. By emergent qualities, I mean intrinsic attributes which are qualitatively distinct from any attributes possessed either by the low-level constituents of physical reality, considered individually, or by any configurations of them that involve relatively small numbers of these constituents, or which have a relatively low level of organization or complexity. The idea is that, at a certain number/density/complexity (or whatever) *threshold*, new qualities emerge which are different in kind from any that are present in sub-threshold phenomena involving these same constituents; and *pari passu* with these new qualities, new behaviour also. One can imagine possessing a dynamical theory which is ostensibly equal to the task of describing the fundamental constituents, and explaining and predicting their behaviour, *up to the threshold*—at which point, however, the theory begins to prove inadequate.

Well, I daresay that something roughly along these lines may be true. Indeed, it is difficult to see how *awareness itself* could be anything other than an emergent phenomenon, in something like the above sense, assuming the truth of materialism. Nor does such emergence threaten to compromise the unity of physical science. Whatever emerged, at and above the associated threshold, would—by hypothesis—have been *latent*, all along, in the low-level constituents. Hence, a complete description of these constituents would have to include reference to dispositional properties, of which the emergent qualities and behaviour constituted a manifestation. If

we assume—as is very plausible—that all dispositions must have a *categorical base* (as the disposition of a key to draw the bolt of a given lock has *shape* as its categorical base), then a description of these constituents need contain no reference to these dispositions as such. It would suffice to cite their intrinsic (non-dispositional) attributes, together with the fundamental laws; a disposition, on the part of any low-level constituent, would hold in virtue of the combination of its intrinsic, categorical attributes and laws which related these attributes to the emergent ones. And incidentally, even if awareness, say, is an emergent phenomenon in the sense just indicated (involving emergent properties and relations), it does not follow that the fundamental low-level constituents need possess any intrinsic, categorical attributes other than those which current physical theory would credit them with—at least, under the conditions prevailing in ordinary physics experiments. Their potential for generating awareness could be a matter of the application of certain currently unknown *laws* to their familiar physical attributes (in which laws, of course, there *would* be an essential reference to the emergent attributes). This fairly elementary point would appear to have escaped those authors who have argued that, if we are made out of electrons, quarks, gluons, and the like, then—given that we are conscious—electrons, quarks, and so forth must themselves be possessed of some sort of primitive proto-consciousness. As I see it, this is a complete *non sequitur*.

So, as I say, emergence in this sense seems to me wholly unobjectionable, philosophically speaking. But, having said that, I doubt very much whether such emergence could, realistically, be expected by itself to offer a solution to the grain problem. For we need to ask: is it really *scientifically* plausible to suppose that the distribution of these emergent qualities would possess any less microstructural complexity than that of the non-emergent ones? Let us go back to our earlier schematic example, involving a circular array of discrete centres of electrical activity in the brain. How, by appealing to emergence, might one explain how this array could present itself to consciousness as an *unbroken* circle? Well, one might suppose that, under the right conditions, such an array would give rise to an emergent field, in the immediately surrounding space, which was continuous, homogeneous, and bounded, in such a way as to match the associated phenomenal presentation, and

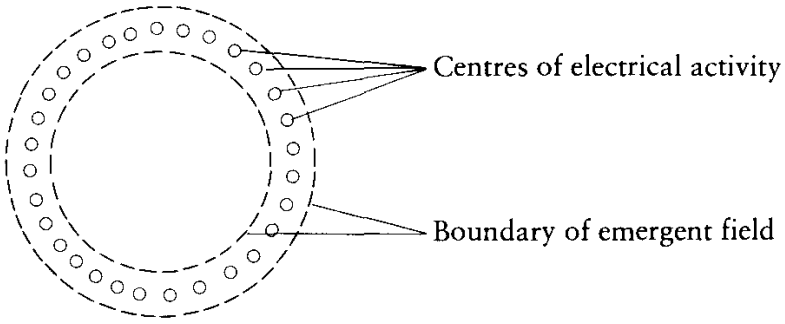

Fig. 12.1. *How a circular array of discrete centres of electrical activity in the brain might give rise to an emergent field*

the innate quality of which was registered in awareness. (See Fig. 12.1.)[15]

In short, we should have to suppose that the microstructural arrangement of the fundamental constituents was capable of giving rise to emergent distributions of qualities which were *truly* smooth and homogeneous, where their 'source' was anything but—in stark contrast to any actual field known to science, and in clear violation of the theoretical demands of quantum mechanics. Well I, for one, simply don't believe it; and I doubt if many people would. Where emergence may indeed come into its own is in accounting for the qualitative diversity that is in evidence within the phenomenal realm: McGinn's problem of how 'technicolour phenomenology [can] arise from soggy grey matter'. But as regards the problem of phenomenal coarse-graining, it seems to me that it has little or nothing to offer. A solution—if such is to be found within the confines of philosophical materialism—must be sought elsewhere.

Before we proceed further, I should make clear just what I take to be required of a solution to the grain problem. The premiss of the problem is that sensory phenomenology belongs, so to speak, to that tip of the neurophysiological iceberg which projects above the surface of awareness. We are to regard it as a part or aspect of the reality of the brain that is directly present to the conscious mind, without benefit of any intervening representation: in awareness, the intrinsic nature of some part or aspect of what is to be

[15] I am here imagining that phenomenal spatial relations, say within the visual field, reflect—indeed, in some sense just *are*—actual spatial relations within the appropriate region of the cerebral cortex. But this is only for the same of concreteness; I do not advance it as a serious hypothesis.

found within the skull stands revealed. From this it follows that the phenomenal objects of sensory awareness must be thought of as somehow *embedded* within that tract of physical reality consisting of the brain and its doings. Assuming this position to be correct, consciousness, at the phenomenal level, can only make selections from the underlying neurophysiology. There is, as we have seen, no possibility of interposing any further stage of reprocessing between awareness and the neurophysiological substrate of its phenomenal objects; for sensory phenomenology is located precisely at the point where the *output* of all processing of sensory data is delivered to the conscious mind. The challenge posed by the grain problem is, therefore, the challenge of showing how *mere selectivity*, as applied to the physical reality of the brain, can yield the form and qualitative content characteristic of sensory phenomenology.

It is often said that philosophers are better at asking questions than at answering them; and I fear that this philosopher is no exception. All that I shall try to do now (and all that space really permits) is to provide a few hints towards a solution.

Underlying the grain problem, it seems to me, are a number of tacit assumptions about the nature of reality and our relationship to it which, though intuitively natural, are philosophically far from compelling. First, I suspect that most people, when it is put to them that in awareness we are immediately aware of some part or aspect of our own brain states, will think that, on this view, the relation between what is going on in one's brain as a whole and the phenomenal contents of awareness must be something like that between a painting and a detail of that painting. But to suppose that is to make the natural but nevertheless unwarranted assumption that the *principle of selection* underlying consciousness must be purely spatial location. There is, *a priori*, no reason to assume that *any* purely *spatial* cordoning of the brain at a given time would capture all *and only* that of which the subject was directly aware. With respect to any spatially defined region, the subject could surely be aware of some but not all aspects or attributes of what lay within it. Secondly and relatedly, there is no good reason to assume that the contents of a given state of awareness correspond to *simultaneous* goings-on in the brain. Indeed, in the context of relativity, no absolute sense can be attached to the notion of spatially separated events being simultaneous. From a

relativistic viewpoint, the brain corresponds to a four-dimensional *world-tube*. And it is as likely as not that what is, so to speak, *given together* in awareness is spread throughout a segment of this world-tube, rather than being confined to a time-slice. In some ways, that would square better, in any case, with the psychological phenomenon of the *specious present*—the fact that, subjectively speaking, states of awareness seem to possess a measure of temporal 'depth'.

These are assumptions relating to us in relation to the world. But there is, thirdly, an assumption about the nature of reality itself which one might wish to question. Kronecker once said, apropos of arithmetic, that the natural numbers were created by God and that everything else is the work of man. In a similar way, it seems to me, people are very apt to suppose that only microstructure is, so to speak, God-given, and that any higher level of structure is, at best, ontologically or metaphysically derivative, and at worst, a mere conceptual artefact. That assumption, in effect, was the basis of our earlier attack on Maxwell's suggestions regarding 'middle-sized structure'. But perhaps, after all, this notion of the ontological primacy of the microstructural is a dogma which should be rejected; perhaps the dilemma on whose horns we attempted to impale Maxwell is a false one. (I shall shortly advance some considerations which suggest that it is.)

None of these observations, however, penetrates to what I conceive to be the real heart of the matter, which is that the grain problem is one manifestation of a more general philosophical puzzle having to do with *subjectivity* and *objectivity*. The world of modern science, it is sometimes said, is a *centreless* world, a world which abstracts from the *point of view* of any particular observer. As Nagel neatly puts it, science is in the business of describing 'the view from nowhere'.[16] Awareness, by contrast, is inescapably centred on a point of view. What is directly present to awareness must, therefore, be conceived as a *perspective* on the brain. I wish to argue that the apparent dissonance between a physiologist's description of brain activity and the contents of our introspective judgements is to be seen, in part, as a consequence of the (relatively) perspective-transcendent character of the former.

If what is true 'subjectively' is true relative to a point of view,

[16] T. Nagel, *The View from Nowhere* (Oxford, 1986).

then the only way of reconciling subjectivity and objectivity is by incorporating *points of view* within one's objective inventory of the world. Any metaphysical theory which does not somehow include points of view in its ontology is to that extent inadequate, as a comprehensive conception of reality. One philosopher who saw this very clearly was Leibniz, who went to the extreme lengths of supposing, in effect, that the universe was entirely composed of points of view—his *monads*.

What I have just said applies as much to physics as it does to metaphysics. Indeed, it is in some sense as much a condition of the explanatory adequacy of a physical theory that one be able to locate, within it, the point of view of the observer, as it is of the practical efficacy of a map that one be able to pinpoint, on the map, one's own position.

In classical physics it was unnecessary to address this requirement explicitly, since the associated conceptual scheme was essentially continuous with that of common sense. In the theory of relativity, however, the requirement is met quite explicitly via the notion of a *frame of reference*. The currently favoured language of space–time and four-vectors would be intuitively unintelligible without the auxiliary notion of an *observer* as situated at the origin of a set of spatial co-ordinates with respect to which he is invariably at rest, with a personal flow of time which corresponds to integration of the space–time *interval* along his own *world-line*. (Einstein sometimes put this very concretely, imagining the observer as carrying around with him an apparatus consisting of three rigid, mutually perpendicular measuring-rods, clutching it at the point where the three rods intersect—the 'origin'—to which is attached an ideal clock.)

Via this notion of a frame of reference, one comes to see how, from the observer's own point of view, space–time comes to be partitioned into the distinct space and time of common sense. Thus, the space–time interval (that is, four-dimensional separation) between two events comes to be decomposed into spatial and temporal intervals corresponding, respectively, to the *projections* of the space–time interval on to the three-dimensional space defined by the spatial co-ordinates (or the set of mutually perpendicular measuring-rods) and the one-dimensional space defined by the time co-ordinate (or the ideal clock). And, in general, a four-vector is decomposed into a three-vector and a scalar

component—the four-momentum, for example, into a three-vector momentum and a scalar energy.

It is tempting to think of an observer, in the context of relativity, as the *concrete embodiment* of a frame of reference (rather than merely as 'carrying around' such a frame, à la Einstein). A description of objects and events *relative to* this frame of reference—couched, therefore, in the language of space and time, rather than of space–time—may then be thought of as corresponding to the observer's perspective: how things are from his own 'subjective' point of view.

The conception of physical science as giving us a centreless account of the world chimes well with its aim, in modern times, of finding things that remain *invariant* with respect to different ways of representing physical reality. (Einstein, it is alleged, originally wanted to call his theory the *theory of invariants*.) This notion of invariance is perhaps the single most powerful idea in modern physics and crops up everywhere—gauge invariance, in field theory, being one of its most prominent recent manifestations. But in particular, it crops up at the foundations of *quantum mechanics*. States of a physical system, in quantum mechanics, correspond to vectors in an abstract space known as *Hilbert space*. And just as the four-dimensional space of relativity can be seen through the eyes of any of an infinity of different *frames of reference*, so the Hilbert space of a quantum-mechanical system can be seen through the eyes of any of an infinity of different so-called *vector bases*. Every quantum-mechanical *observable*, that is to say, question one can ask of a quantum-mechanical system, susceptible of being answered by a suitable measurement—or more generally, every set of *compatible* observables, that is, questions capable of being simultaneously answered—corresponds to a vector basis for Hilbert space, known as the *eigenbasis* of the corresponding observable or set of observables. A set of observables, via its *eigenbasis*, defines a co-ordinate system for Hilbert space, just as a frame of reference defines a co-ordinate system for space–time. (The key things that remain invariant with respect to different bases are the respective probabilities of getting various outcomes, when carrying out measurements on the quantum-mechanical system in question.)

Quantum mechanics was discovered independently, in the mid-1920s, by Heisenberg and Schrödinger. But so different, in their

mathematical formulation, were Heisenberg's *matrix mechanics* and Schrödinger's *wave mechanics*, that they were at first thought to be distinct, rival theories. Only subsequently were they found (by Schrödinger himself, in the first instance) to be essentially equivalent, the difference in form being due, in large part, to different choices of basis. Roughly speaking, Heisenberg chose the eigenbasis corresponding to the energy observable, and Schrödinger the eigenbasis corresponding to the position observable.

That said, I am now in a position to convey the essence of my own favoured approach to the grain problem. (Again, space does not permit more than a very approximate and compressed rendering.) First, the brain, I suggest, may legitimately be regarded as a quantum-mechanical system. (There is nothing in the formalism of quantum mechanics that prevents it from being applied to macroscopic systems.) As with most complex quantum-mechanical systems, the brain may be conceptually decomposed (doubtless, in several alternative ways) into *subsystems*, which can be treated as quantum-mechanical systems in their own right. One of these subsystems I take to coincide with the neurophysiological substrate of conscious mental functioning. (This dividing-up of the brain into subsystems need not, as remarked earlier, correspond to anything that would seem intuitively at all natural; nor need the subsystem in question correspond to what would ordinarily be thought of as a *part* of the brain. The dividing-up would not, in general, be a *spatial*, or even a *spatio-temporal* division, so much as a partitioning of the *degrees of freedom* of the larger brain system—the distinct ways in which it can change state or store energy.)

Anyway, there is, I take it, such a brain subsystem. And from the point of view of consciousness, I contend, there is (at any given time, at least) a preferred set of compatible observables on that system. The conscious observer views his or her own brain through the eyes, so to speak, of this preferred set of observables, much as the observer, in relativity, views the world through the eyes of his own frame of reference. Indeed, just as, in relativity, the observer can, in a sense, be regarded as a concrete embodiment of a particular frame of reference, so, I suggest, may a conscious subject be thought of as, in some sense, the concrete embodiment of a set of compatible observables. Every quantum-mechanical observable has a so-called *spectrum of eigenvalues*, associated,

respectively, with the eigenvectors comprising its eigenbasis; these are numbers corresponding to the possible results of measuring the observable in question. If we consider a *set* of observables, then their spectra can themselves be thought of as co-ordinate axes, jointly defining a further abstract space. And a value of each co-ordinate, corresponding to an eigenvalue of each observable in the set, will define a point or vector in this abstract space. When the set of observables is the set preferred by, or embodied in, consciousness, then this space may be equated with phenomenal or experiential space: points or vectors in the space correspond to distinct possible states of awareness. And the various *qualia* spaces of sense-perception—colour space, for example—are simply subspaces of this abstract space; specific *qualia*, as they figure in awareness, represent points or regions of such spaces encountered, so to speak, in the flesh. It is precisely here that the intrinsic character of the concrete reality that the abstract mathematical formalism of quantum mechanics purports to describe makes itself manifest.

But how does all this address the problem of how awareness is able to gloss over the complex microstructure which presumably underlies the phenomenal contents of any experience? Well quite simply, there are, in quantum mechanics, no observables, or sets thereof, which are *a priori* privileged. In particular, there is, in terms of quantum-mechanical observables, no rock-bottom level of structure to be discerned in the world. In quantum field theory, no sense can be attached, for example, to the notion of measuring the values of the field variables at a precise point—only their average values over some finite spatio-temporal region (which one can make as small as one wishes); indeed, no sense can be attached to their *possessing* a precise value at any precise point. (No more, in elementary quantum mechanics, can an electron be said to have, let alone be measured as having, a precisely defined position or momentum.) In quantum mechanics there is a sense in which all observables, and in particular observables corresponding to every level of structure, are to be regarded as equal in the sight of God, as are different frames of reference, relativistically conceived.[17] As I intimated earlier, quantum mechanics seems to be

[17] For the benefit of those familiar with quantum mechanics, let me say that I am, of course, glossing over the distinction between so-called *maximal* and *non-maximal* (or *degenerate*) observables, or sets thereof. (A maximal observable, or set of observables, is one corresponding to a measurement, or set of simultaneous

telling us that it is a classical prejudice to suppose that the world is not *intrinsically* structured at anything but the level of elementary particles, and their actions and interactions.

According to this approach, then, the apparent dissonance between brain activity, as seen through the eyes (and concepts) of the neurophysiologist, on the one hand, and the conscious subject, on the other, is to be attributed to three distinct considerations. First, this brain activity is revealed to the awareness of the corresponding conscious subject—as it is not to the probings of the neurophysiologist—*as it is in itself* (albeit only from a certain point of view). Second, introspective awareness is focused on a *subsystem* of the brain, selected according to principles that, from the standpoint of physiology, would seem very unnatural. And finally, the contents of consciousness correspond to eigenvalues of a set of *observables* which, again, are distinct from anything that the physiologist is likely to settle on: the dissonance between the subject's view, and that of the physiologist, might be conceived as analogous to that between, say, Schrödinger's wave mechanics and Heisenberg's matrix mechanics. Thinking in terms of co-ordinate systems for the Hilbert space of the relevant brain system, it is as though the co-ordinate system of the conscious subject were, so to speak, rotated with respect to that of the external observer.[18]

The state of a physical system—on the view that I am proposing —might be compared to a block of wood, distinct cross-sections

measurements, which yields a state of *maximum information* about the system in question—one that cannot be improved upon by the performance of further measurements.) In case someone thinks that maximal observables, or maximal sets of compatible observables, are privileged with respect to non-maximal ones, in a way that vitiates my argument, it should be pointed out that one could imagine the space of possible states of awareness of the conscious observer being generated, so to speak, in two stages. Any non-maximal set of compatible observables can, after all, be turned into a maximal set simply by adding observables to the original set. So suppose, to begin with, that there is (from the perspective of consciousness, though not of the world) a preferred maximal set of compatible observables (having the requisite non-maximal set as a subset). The spectra of eigenvalues of the observables in the set could then be thought of as co-ordinate axes, defining a state space, with respect to which the range of possible states of awareness could then be regarded as constituting a preferred subspace.

[18] Here I have been able to do no more than sketch the bare bones of the theory I favour. In *Mind, Brain and the Quantum*, I develop these ideas in far greater detail, and also, for the sake of those unversed in modern physics, provide an elementary account of quantum mechanics itself.

of which can reveal strikingly different sorts of patterns, depending on the angle at which it is sliced: concentric rings at one extreme, roughly parallel, gently undulating lines at the other. Though the analogy is very imperfect, one might think of the neurophysiologist, and the conscious subject in introspection, as likewise being confronted, so to speak, with different 'cross-sections' of what are in fact the same brain states. My claim is that, by appealing to the quantum-mechanical concept of an observable, we can render it intelligible, as with the grain of the wood, that a common underlying structure should manifest itself in superficially very different ways. On the side of introspection, moreover, such a conception removes the need to appeal to any inner representation, distinct from the state itself. For to be directly acquainted with a 'cross-section' of something is *a fortiori* to be directly acquainted with the thing itself, not merely some cognitive surrogate of it—in spite of the fact that what is thereby revealed to consciousness is revealed only under a certain aspect.

What, then, finally, is consciousness telling us about the nature of physical reality? Well, first (assuming materialism to be true), that there is more to matter than meets the physicist's eye. For there is nothing in the physicist's account of the world to explain why there should *exist* conscious points of view—why the world should contain such concrete embodiments of sets of quantum-mechanical observables. Thus we are in a position to know *a priori* that something like superstring theory, whatever its other merits, cannot literally be a theory of everything—since there is nothing in the theory that could, even in principle, explain how matter, suitably combined, is able to generate awareness. But on the positive side, it follows from what I have been saying that our states of awareness, corresponding as they do, on my account, to sequences of eigenvalues of brain observables, are providing us with the answers to specific questions concerning the intrinsic nature of a corner of the physical world—something that (as Russell rightly insisted) can never be revealed in ordinary measurement or observation. For our own awareness, so I have been urging, embodies a preferred set of observables, which in turn amounts to saying that its contents, at any given time, embody the answers to a set of questions about the state (the intrinsic state) of the underlying brain system. Sadly, however, we here find ourselves in a predicament akin to that of the characters in *The Hitch Hiker's Guide to*

the Galaxy, on being told that the answer to life, the universe, and everything was 42. We know the *answers* to these questions, in a way that a scientist, merely by examining our brains from without, never could. But unfortunately, we have, as yet, no idea what the questions are!