Problems in Philosophy

The Limits of Inquiry

Colin McGinn

BLACKWELL
Oxford UK & Cambridge USA

gave at Princeton I enjoyed the sympathetic criticism of, among others, Alex Burn, Fiona Cowie and David Sosa. My greatest debt is to Noam Chomsky – for the stimulus provided by his work, for much encouraging and probing correspondence, and for detailed comments on the penultimate draft of the book. I would also like to thank my department at Rutgers for giving me leave and for the wholesome intellectual atmosphere there.

Colin McGinn
New York

1

Philosophical Perplexity

I Transcendental Naturalism

Philosophical theses can sometimes be assented to, but often they can expect only to be taken seriously. We may hope to find sufficient reason actually to believe a philosophical proposition, but often enough the best we can do is get ourselves into a position to regard the proposition with respect. We rate it a contender. The set of such propositions constitutes the range of options we think *may* be true: they cannot be ruled out and there are considerations that speak in their favour. We might, on occasion, have reasons – more or less indirect, or large-scale – that make us suspect that a certain hypothesis *must* be true, without being able to show *how* this is so. The hypothesis may not engage our beliefs in the simple way other hypotheses do, in science or common sense, yet we find ourselves irresistably drawn to it: something like this, we feel, *has* to be so, if only because it is preferable to its rivals. A good deal of philosophical debate consists in persuading others to take seriously a hypothesis one has come to find attractive for reasons that defy summary statement or straightforward demonstration; one tries to exhibit the virtues of the hypothesis, especially as compared to the available alternatives. Or, failing even that, one argues that the hypothesis cannot be excluded. One limits oneself, modestly, to soliciting philosophical respect, recognizing that philosophical belief is too much to expect – not to speak of philosophical certainty. And this shows something significant about the nature of philosophy – about the

epistemology of philosophical inquiry. The relation between evidence (argument) and truth is very often not close enough to permit full-blown assent. Hence the magnitude and intractability of much philosophical disagreement.

My aim in this book is to try out a very general hypothesis. The attitude I intend to produce towards the hypothesis is mere respect; if the reader ends up believing it, that is his or her own business. In the nature of the case, indeed, it is a hypothesis which does not admit of the kind of demonstration we naturally demand for hypotheses of its general form. My claim will be that the hypothesis may be true, and that much would make sense if it were. I shall proceed as follows. In this chapter I shall set out, in a preliminary way, what the hypothesis says, explain the conception of philosophical problems it entails, and sketch the geography of philosophical debate it predicts. This will serve to introduce the basic perspective and the conceptual apparatus needed to articulate it. In the next six chapters I shall apply the hypothesis to a number of specific issues, treating the problem of consciousness as a philosophical paradigm, testing the plausibility of the hypothesis and drawing out its consequences: here is where I hope to sow the seeds of philosophical respect. Chapter 8 will draw this material together and explore the claims of human reason to be the route to philosophical truth. The following chapter will briefly compare the resulting view of philosophy with some other views, arguing for its superiority. Throughout we shall be operating at a high altitude, covering a lot of ground in a rather abstract and tendentious fashion, stressing metaphilosophical themes. The discussion will often be distressingly speculative and schematic, even by philosophical standards.

What, then, is this putatively reputable hypothesis? Very simply it can be put as follows: philosophical perplexities arise in us because of definite inherent limitations on our epistemic faculties, not because philosophical questions concern entities or facts that are intrinsically problematic or peculiar or dubious. Philosophy is an attempt to get outside the constitutive structure of our minds. Reality itself is everywhere flatly natural, but because of our cognitive limits we are unable to make good on this general ontological principle. Our epistemic architecture obstructs knowledge of the real nature of the objective world. I shall call this

thesis *transcendental naturalism*, TN for short. Let us now try to sharpen TN up a bit, explaining its motivation and key concepts, preferably without the aid of metaphor.

We need, first, to make some basic distinctions among questions that may interest the enquiring mind. Four sorts of question may confront a particular type of cognitive being B: problems, mysteries, illusions, issues.[1] A *problem* is a question to which B can in principle find the answer, and is perhaps designed so to do, for biological or other reasons; or at least is of such a *type* as B can answer. Everyday life and much of science consists of solving problems – questions that fall within our cognitive bounds. A *mystery* is a question that does not differ from a problem in point of the naturalness of its subject-matter, but only in respect of the contingent cognitive capacities that B possesses: the mystery is a mystery *for that being*. An *illusion* is (or arises from) some kind of pseudo-question, or a question that is so formulated as to suggest an answer of a kind that does not objectively exist. An illusory question is not to be confused with a mysterious one, which latter reflects ill on B, not on the question. An *issue* is a question, typically of a normative character, about which B creatures may dispute, and with respect to which no scientific theory is suited as an answer – questions of ethics and politics, say.

Note, what is vital, that the categories of problem and mystery are defined in a relative way – the division turns upon the specific cognitive make-up of the class of thinking creatures we are considering.[2] In principle, two sorts of creature might invert each other's characteristic division of questions into problems and mysteries, depending on their epistemic talents and endowments. Indeed, the mystery class for one sort of creature might be *innately* soluble by the other, and this be manifested at an early age too. Let us say that such creatures differ in the 'cognitive space' through which their minds can move, rather as there exist species-specific differences in the motor spaces through which creatures can physically move, given their natural constitution (birds and fish, say). Then the idea of a mystery is simply the idea of a question that happens to fall outside a given creature's cognitive space. It is analogous to the idea of items that lie outside of a creature's phenomenal or perceptual or affective space – sensations it cannot

feel, properties it cannot perceive, emotions it cannot experience. If we suppose that creatures possess 'organs' that define these spaces, then mysteries are questions for which the given creature lacks the requisite intellectual organ(s).[3] The totality of these spaces would constitute the mental horizon of the type of creature in question.

We can also capture the underlying idea here by means of a counterfactual about a given creature: take a question that is a mere problem for the creature (say, a human being) and imagine the creature to have the relevant problem-solving capacities removed, so that the question is no longer answerable by that creature. In the counterfactual case, then, there is no intrinsic change in the ontological status of the topic of the question – we have simply moved to a situation in which the relevant creature now lacks the epistemic capacities to comprehend that topic. By hypothesis, the change is purely epistemic. Steam engines do not turn occult when the possible world in which they exist happens to lack any creatures with the mental capacity to understand their workings. And the converse shift, from mystery to problem, likewise involves no sudden access of ontological purity. This is really no more than to insist on the epistemic character of the distinction.

According to TN about a certain question Q with respect to a being B, the subject-matter of Q has three properties: (i) reality, (ii) naturalness, and (iii) epistemic inaccessibility to B. Q does not harbour an illusion (hence (i)), nor does it refer to entities or properties that are intrinsically non-natural (hence (ii)), yet the answer to Q is beyond the capacities of B creatures (hence (iii)). Thus TN contrasts, not only with illusion theses about Q, but also with three other positions, as follows. Immanent naturalism takes all genuine questions to have answers in the space of theories accessible to B, and it is comprehensively naturalistic. Immanent non-naturalism accepts an ontological bifurcation into the natural and the non-natural but insists on the comprehensibility to B of both sides of the bifurcation, perhaps in virtue of supernatural capacities on the part of B. Transcendent non-naturalism asserts that some questions invoke facts that are both supernatural and beyond the capacity of B to comprehend. TN, for its part, makes no ontological division into the natural and the non-natural, but

it is happy with an epistemological division into the answerable and the unanswerable. In fact, it is truer to the intentions of TN to avoid any positive use of the term 'natural', since this implies a well-defined distinction between the natural and the non-natural; better to think of TN as opposing the very idea of the non- or super-natural altogether. TN is anti-non-naturalistic: it is the negative thesis that there is no sense to be made of the ontological category of the non-natural. It regards such notions as distorted reflections of epistemic quandaries, not as signifying meaningful objective categories.[4] (This should become clearer later when we see TN in operation.)

Plainly TN accepts a strong form of realism; in particular, it accepts realism about the *nature* of the things that cognitive beings think and talk about. While we may be able to refer to certain things, there is no guarantee that we shall be able to develop adequate theories of these things. Put differently, the correct theory of what is referred to, conceived as a set of propositions detailing the nature of those referents, may not belong in the space of theories accessible to the beings under consideration – including human thinkers. So, for TN, there may exist facts about the world that are inaccessible to thinking creatures such as ourselves. Reality is under no epistemic constraint.[5]

TN, as I am defining it, incorporates a double naturalism (or anti-non-naturalism – I shall drop this periphrasis from now on): both about reality and about our knowledge of it. The natural world can transcend our knowledge of it precisely because our knowledge is a natural fact about us, in relation to that world. It is a general property of evolved organisms, such as ourselves, to exhibit areas of cognitive weakness or incapacity, resulting from our biological constitution; so it is entirely reasonable to expect naturally based limits to human understanding. We are not gods, cognitively speaking. A creature's mental powers are things *in* the natural world, with a natural origin, function and structure, and there is no necessity that this part of the world should be capable of taking in the rest. The 'transcendent' component of TN simply gives expression to this naturalism about the mind.

The transcendence envisaged by TN can take stronger or weaker forms, and many versions of it can be formulated. For my purposes, the most pertinent distinction to introduce is between *bias*

theses and *closure* theses. A bias thesis holds that the faculties of B are skewed away from certain questions, possibly because they are skewed towards others. More precisely, and adopting a modular conception of cognitive capacity, we think of B's epistemic potential as the sum of its several cognitive modules – special-purpose domain-specific systems – and these modules will have inbuilt principles biasing them away from dealing with certain types of question. For example, the human language module is negatively biased with respect to certain definable languages, and it is no use at all in developing other sorts of knowledge.[6] Within a given type of creature the bias of a module M1 may be compensated for by a distinct module M2, so that the creature is able in principle to know what M1 prohibits – as we might laboriously learn languages for which our given language module is unsuited by exploiting our general capacity for theory construction. But it is also possible that the biases of the totality of the creature's modules completely rule out acquiring certain sorts of knowledge. In the former kind of case the creature will experience considerable subjective difficulty in acquiring the knowledge in question, no matter what objective properties the domain in question possesses. In the latter kind of case there will be no escape from the cognitive bias and ignorance will be irremediable: the bias will lead to closure. If we picture the mind as analogous to a Swiss army knife, where each gadget corresponds to a cognitive faculty, then there will be tasks for which no gadget on the model of knife in question can do the job and tasks for which a gadget designed for one kind of job can be pressed into service in executing another. Given the truth of a negative bias thesis with respect to a certain subject-matter, of either of these two kinds, we would expect all the symptoms of immense difficulty, both behavioural and subjective, combined with no special reason to suppose that we have entered the realm of the inherently inscrutable or impossibly complex. And this may alter our conception of the nature of that subject-matter. Bias does not, then, entail closure, though it may well underlie it in cases for which closure holds. Neither does closure entail bias – or not obviously. A question may be necessarily unanswerable for a creature consistently with a *tabula rasa* conception of that creature's mind: that is, we can view cognitive capacity as non-modular and

undifferentiated, as a 'general-purpose learning machine', and still hold that certain questions exceed the limits of such a mind, say by dint of sheer complexity.[7] Bias theses entail a particular view of cognitive structure; closure theses imply merely the existence of cognitive limits.

Now if our class of questions is that of philosophy, we can formulate TN either in terms of closure or bias, or closure in virtue of bias. A bias thesis by itself is weaker than a closure thesis in that it allows the possibility of solving philosophical problems; what it claims, however, is that there is a mismatch between module and question in the philosophical case. Hence the difficulty of arriving at philosophical knowledge: the epistemic characteristics of philosophical questions result from a systematic bias away from the subject-matter of those questions. In what follows much of the discussion will be neutral between these two interpretations of TN, though it will become apparent that I am inclined towards a thesis of modular closure.

A fanciful parable may help to bring out the import of TN for the nature of philosophical questions. Imagine a race of intelligent beings who suffer from the following cognitive deficit: they can form no conception of what it would be for material objects to have atomic structure – in particular, for objects to contain internal spatial interstices. They cannot help but think of objects as having a continuous structure, with no empty space inside them. They understand the idea of gaps between macroscopic objects, but a quirk of their conceptual system prevents them forming even so much as the idea of gaps within objects. Now suppose these thinkers ask themselves, as well they might, how it is possible to divide an object into parts, or how an object can be compressed. And let us assume, for the sake of argument, that these questions have a unique solution in atomic theory: objects are divisible or compressible in virtue of the gaps that separate their smaller parts. Then, given all this, we can say that these questions are unanswerable by the continuous-thinkers – though answerable by atomic-thinkers such as ourselves. In the sense explained earlier, these questions belong in the class of mysteries for them, because they have a bias in favour of continuity that generates closure with respect to questions whose answers require knowledge of (simple) atomic theory.

Similar examples could be given by imagining creatures who cannot form concepts of three-dimensional space, or who cannot form the concept of a negative number, or whose notion of causality is limited to mental causation, or who can only think about the present, and so forth. In all these cases we can envisage questions that require conceptual and theoretical resources that exceed the contingent limits of the creatures in question. And the TN hypothesis would (virtually by stipulation) be true of them. Specifically, their mysteries would have the appearance of deep philosophical conundrums, analogous to our own philosophical puzzlement – that, at any rate, is what TN about human philosophy is in the business of claiming. *We* can see that the division problem, say, is just a problem in low-level science, not something that calls into question the entire ontology of material objects or requires the postulation of miracles or inexplicable brute facts. The special hardness they find in the problem is a reflection of a definite lack on their part, with no ontological implications. This is what TN says is the case with respect to the philosophical problems that trouble us. TN locates philosophical profundity in the specific cognitive deficits we suffer.[8]

II The Nature of Philosophy

Let us remind ourselves of how philosophical perplexities typically arise, and of the form they are apt to take. Common sense commits itself to various assertions about the world, including the mind. We acquire these ordinary beliefs at an early age and we take them for granted in everyday life; they probably have an innate basis and belong to a specialized component of our cognitive equipment. Then, because we are also self-reflective creatures, we turn back on our commonsense assumptions and find them to be more puzzling and problematic than we had bargained for. The concepts we habitually employ raise the kinds of disturbing questions we call 'philosophical'. A characteristic expression of this puzzlement asks how what we had hitherto taken for granted is actually so much as *possible*. Let C be a concept that provokes this kind of question: then the philosophy of C will concern itself with whether C-truths, commonly taken for granted,

are, in the light of certain considerations, really capable of being true at all, and if so in what their truth might consist. C-propositions seemed to work perfectly well in practical contexts but, upon examination, they present *prima facie* impossibilities. The putative C-truths clash (it appears) with certain other beliefs we hold about the world, and the question is how to retain both or decide what to give up. We thus strive to understand the nature of C-facts in such a way that it becomes clear that the world can contain such facts. The simplest form of philosophical perplexity is accordingly expressed by the question 'What *is* X?'. Not that every philosophical question assumes this form; nor that philosophy is exclusively concerned with how-possible questions: but a substantial core of it traces back to this kind of perplexity (as we shall see in detail in later chapters). And it is a perplexity of a peculiarly knotty kind, generating intimations of ultimate mystery, a dazed sensation where knowledge ought to be.

A notable feature of these philosophical problems is that they seem to be about things in the world and yet are not answerable by empirically investigating those worldly things: that, at least, is how they naively present themselves. So they are like scientific questions in one way but unlike them in another (of course science itself may also raise philosophical questions). This can seem puzzling: if they are about worldly phenomena, why won't they yield to world-oriented investigation? Indeed, this combination of characteristics has seemed so puzzling to many (most) philosophers that they have revised the initial appearances: either the questions are not about the world after all or they really are answerable empirically. Thus we have the two dominant metaphilosophies – two conceptions of the proper subject-matter of philosophy, of the type of truth it endeavours to discover, of the right method to follow in discovering this truth. In effect, the two metaphilosophies differ in respect of the kinds of human faculty they take to be appropriate in arriving at philosophical knowledge, and hence in where we should look for philosophical enlightenment.

The two approaches are, familiarly enough: (a) the view that philosophical questions are (ultimately) empirical or scientific and (b) the view that they are (upon reflection) analytic or conceptual. The former view takes philosophy to be continuous with

extant science, so that its questions become incorporated into science in the fullness of time (or else they are declared meaningless). Accordingly, the human faculties to use in doing philosophy are the same as those we use in empirical science: our powers of perceptual observation and our talent for empirically controlled theory construction. Philosophy is just the outer edge of empirical inquiry. In opposition to this we have the school, dominant for most of the present century, that radically distinguishes philosophy from science. This school regards philosophical inquiry as conceptual in topic and method: we are to answer our how-possible questions by elucidating the concepts that occur in our C-truths, using our faculty of self-reflection; or again, we must ruminate on the language we bring to bear on the sector of thought at issue. Hence the conception of philosophy that prompts what is sometimes called 'analytic philosophy' – a belief in the problem-solving potential of our ordinary concepts, once they are scrutinized aright. Philosophical knowledge will thus issue from the same human faculty that enables us to know (e.g.) that bachelors are unmarried males. This view embodies a principled optimism about the capacity of our present conceptual scheme to resolve questions about its own presuppositions; it is conceptually conservative in a way the contrasting empirical conception is not. The nature of the facts that trouble us is implicit in our ordinary C-propositions, waiting to be excavated a priori; so really the questions are about the concepts themselves, at least in the sense that we need not look beyond them.[9]

Where does TN stand in relation to these two standard positions? I shall answer this now in a sketchy and preliminary way, not intended as a defence or full account, but rather to set up the issues for later discussion. First, TN accepts at face-value that philosophical questions are about the world but are not soluble empirically: they concern the real objective nature of concept-independent phenomena, but we cannot answer them by means of empirical enquiry into those phenomena. So TN differs from the other two views while sharing an aspect of each. It accepts, with the empirical view, that philosophy is ontologically continuous with science, but it denies that this corresponds to any epistemic continuity, since we are not cognitively equipped to solve philosophical problems. On the other hand, it agrees with

the analytic view in rejecting the idea that philosophical puzzlement will yield to empirical enquiry, but it denies that this is because concepts are our real concern. From TN's point of view, these rival metaphilosophies distortedly reflect the true epistemological predicament we are in: namely, that we can formulate questions about the world that we lack the faculties to answer. Understandably, then, we deny that they are about the world or insist that future science holds the key, unimaginable as this may seem. TN is pessimistic about the power of human empirical enquiry to solve philosophical problems, not because these problems involve an ontological shift from the world to our representations of it, but rather because we lack the means to reveal the objective nature of the things we refer to. Revising the philosophical appearances, as the other two views do, stems, for TN, from a reluctance to acknowledge our cognitive limits – from a kind of reflex optimism about human knowledge.

By implication, TN casts a sceptical eye on certain paradigms for philosophical enquiry that have been historically influential. Once a certain method of enquiry achieves notable results in other domains, while philosophy appears embarrassingly enmired, it is natural to hope that the way forward consists in following that method in the case of philosophy. Thus the remarkable scientific advances begun in the seventeenth century might well suggest to the hopeful enquirer that philosophical questions will succumb to essentially the same methods – or else reveal themselves not to be genuine questions at all. But TN reminds us that these methods might have a strictly defined sphere of potential success. Natural science is a product of the human mind, with its inbuilt principles and limits, and there is no good reason to believe that every question about Nature can be answered by a mind so structured and employed. Philosophy, in particular, might require styles of thought and methods of enquiry that lie outside the bounds of our capacity for empirical science. And, of course, on the face of it philosophical problems are not soluble by scientific methods. TN provides a perspective from which this epistemic discontinuity becomes intelligible.

A different paradigm, going further back, has been supplied by the formal sciences. Here I mean to include rational ethics as well as logic and mathematics – the non-empirical areas of human

knowledge. The natural thought here is that philosophy should be assimilated to these subjects, on account of its non-empirical character. So we have the idea that the methods of philosophy include dialogue and argument, thesis and counter-example, proof and intuition, analysis and axiomatization. But TN questions whether the human faculties employed in those activities are appropriate for philosophy, since it cannot simply be assumed that they are, in view of the inappropriateness of empirical methods. And how is this paradigm to be squared with the apparent fact that so much of philosophy is concerned with natural phenomena, not abstract or formal matters? TN warns that such paradigms be approached with extreme caution; certainly we should not cling to them simply because no other set of methods suggests itself – for there may *be* no other method available to us. A being who could answer our philosophical questions with comparative ease might use methods and faculties that are radically disjoint from any that we possess; they may even be inconceivable by us, even in rough outline. In short, we should examine putative paradigms for philosophical knowledge on their merits and not be credulously seduced by the lack of anything better – since TN is not to be ruled out. I myself would say that the plain incredibility of the standard proposals for an epistemology of philosophy ought to make us view the TN hypothesis with some seriousness.

Any account of the epistemology of philosophy ought to have something to say about the chronic lack of progress that seems endemic to the subject, compared to other intellectual pursuits. And metaphilosophies can be evaluated according to their ability to explain this lack of steady advance. Again, without taking up the question in detail, let me just state where TN stands as compared with other views. The two standard conceptions encounter obvious *prima facie* difficulties over this question. The empirical view cannot point to the kinds of scientific advance enjoyed by its preferred paradigms: philosophy does not look much like flourishing science. There ought to be more advance than there is, under this view. The analytic view must face the question of why our concepts are so opaque and inaccessible to us: if all we have to do is spell out what our ordinary notions involve, why has it proved so hard to push the subject forward

– it ought to be easy! So, again, there should be more advance than we observe. In response to these challenges adherents offer extenuating explanations. Perhaps there is more progress than there seems, since science keeps slicing off parts of philosophy, leaving a standing residue of not-yet-solved questions.[10] Perhaps our efforts to articulate the content of our concepts are thwarted by certain temptations, misleading analogies, linguistic bewitchments, and so forth. Yet other views hold that the questions of philosophy are meaningless and hence trivially unanswerable, or that they concern matters of such intrinsic subtlety and profundity that lack of progress is hardly surprising. TN has a simple and straightforward explanation to offer: our minds are not cognitively tuned to these problems. This is, as it were, just a piece of bad luck on our part, analogous to the lack of a language module in the brain of a dog. We make so little progress in philosophy for the same kind of reason we make so little progress in unassisted flying: we lack the requisite equipment. We have gaps in our cognitive skills as we have gaps in our motor skills – though in both cases we can see what we are missing and feel the resulting frustrations. That, for TN, is the kind of thing the hardness of philosophy consists in: not bewitchment by the surface forms of language, not deep implicitness in our conceptual scheme, not sheer meaninglessness, not objective complexity or intricacy or non-naturalness. None of these explanations would suit the case of the human inability to fly unaided, or to perceive ultraviolet light, or to hold ten thousand items in short-term memory; and TN holds that the kind of thing that explains these deficiencies is the kind of thing that makes us so poor at solving our philosophical problems.

Given this type of explanation, TN has a particular, and deflationary, account of our sense of philosophical depth. It is often supposed, if only tacitly, that the depth that philosophical questions appear to have is a reflection of some intrinsic feature of their subject-matter, difficult though it is to identify that feature in any illuminating way. TN opposes this tendency: it credits us with a propensity to commit a projective fallacy when we encounter a philosophical problem. We spread our own epistemic shortcomings onto the phenomena that perplex us, so raising the spectre of occult ontology. It then seems to us that the natural world

contains metaphysical oddities, things whose very possibility comes into doubt. And hence philosophy takes on a semblance of special depth, as if it has to wrestle with facts of a peculiarly refractory nature, this impenetrability being grounded in objective reality. Philosophy is then apt to become a debate about whether there really are such facts after all. But TN counsels us to recognize this act of projection for what it is: we are mistaking a cognitive deficit on our part for an objective feature of what we are trying to understand. For a species gifted where we are deficient philosophical questions might have no more depth than we find in elementary geometry. In a sense, then, the depth is illusory, at least as a non-relative trait of the philosophical subject-matter. The predicate 'is philosophically deep', as applied to some worldly phenomenon, signifies a mind-dependent property, rather as the predicate 'is invisible at night' signifies a relation to some specific type of visual system. Indeed, the very concept of the philosophical, for TN, involves the idea of a problem that presents itself as a mystery, relative to some set of faculties of understanding. Creatures who understand our philosophical subject-matter with the ease we find in learning the simple properties of space and matter (say) would not reserve a special category of question labelled 'philosophy'. The word connotes a special kind of intellectual cramp or aura, and they are free of all that.

So far I have set out the TN view in an introductory way, trying to develop a feel for what it says; I have yet to offer any defence of it – a task to be undertaken in later chapters. But before I begin that task I need to introduce some further apparatus to be used in conjunction with TN. Again, we shall be proceeding abstractly, deferring applications till later.

III Philosophical Geography

Philosophical debates tend to assume a characteristic pattern, with an array of options staked out and variously occupied. It will be useful for our purposes to identify this pattern, so that we can apply it in particular cases with TN in mind. The pattern may not always be clearly inscribed on the surface of debate, but I think it almost invariably lurks beneath. To this end, then, I

shall introduce the DIME shape – the shape of the philosophical landscape as it is configured by the underlying how-possible questions. Consider a philosophically problematic concept C, with respect to which we wonder how it is possible that C should apply to the world; so we are going to need to do some philosophy on C if we are to understand what it is all about. Then the DIME shape displays four types of philosophical position that might be taken with respect to C, as follows.

D corresponds to the idea that C must be domesticated, demythologized, defanged, demoted, dessicated. Taken at face-value C presents large problems of understanding and integration, so in order to secure its objective possibility we need to redescribe it in some way. Simple reduction to a relatively unproblematic set of concepts is the standard manouevre here. The thought behind D is that C presents its referent in a misleading and inflated way, exaggerating its ontological uniqueness, so that we need to prune its pretensions somewhat. We must make C-facts humdrum and hence feasible. D may then incorporate an error thesis about C, either for an aspect of the concept itself or for the imaginative flights it provokes in us. However, the intentions of the D adherent are not to expel C-truths but to retain them, after some necessary wing-clipping. The position is that C-facts are really not so remarkable after all, upon a closer inspection of their content. Once unmasked, there is room for them at Nature's inn. They may seem set apart, but actually they are nothing but such-and-such innocent thing in disguise.

The I position is that C-facts are irreducible and indefinable and inexplicable, and we should cultivate an attitude of insouciance towards them. C-propositions state brute facts for which no explanation can be given and for which none should be sought; they are what they are and not another thing. D-style reductionism stems from misplaced monism, obsessive unification. We need to rid ourselves of the compulsion to oversimplify the world, to level it ontologically, and instead relax, indolently, into the *sui generis* variety of our given conceptual scheme. After all, it was functioning perfectly well before we started to fret over it. Apparent clashes between C and other beliefs we hold must result from mistaken philosophical theories or some slip in our thinking. Reality contains C-facts, and there's an end to it. Explanations

must terminate at some point. C-facts are possible because they *are so*, and we can see that they are, irreducibly.

M stands for magical, miraculous, mystical . . . mad. The M believer accepts C-facts at face-value, unlike the D theorist, but he cannot simply take them as inexplicable, like the I adherent – he wants some account of their nature or basis. He seeks a larger picture of the world – a metaphysics – within which C-facts find an intelligible place. He is mightily impressed with C-facts, but he doubts they can be fitted into the natural order. His view is that the world is a stranger place than some people are prepared to admit. To make sense of it we need to invoke God or some equivalent supernatural entity of force. Reality thus includes more than the natural world; and it must, on pain of having no explanation of the facts. Only an M ontology can properly accommodate the data. C-facts are possible only because the world of science is not the only world there is. There are traces of the divine embedded in C. Indeed, the M believer is often to be found trying to establish the truth of a supernatural metaphysics on the basis of C-facts, by deduction or inference to the best explanation: for atoms in the void could never generate such remarkable facts. And even if the acceptance of an M position is less candid and enthusiastic than this, it can sometimes seem that nothing else will suffice: the philosopher may find himself driven in an M direction, perhaps concealing this move from himself, by the extreme difficulty of producing any coherent naturalistic account of the phenomenon in question. Indeed, we might say that the threat of M is partly definitive of a philosophical problem.

E is for elimination, ejection, extrusion. The E proponent despairs of domestication, balks at irreducibility, and scoffs at magic. His position is that C-facts look impossible because that is what they are: they are either prescientific remnants or logical absurdities of some sort. The entire C-ontology is an enormous illusion. C-talk should thus be banned, at least in serious contexts. The reason putative C-facts give us so much theoretical trouble is that we are trying to make sense of the non-existent. At best we might fashion a surrogate for them, to occupy their practical place, but in sober truth C-concepts have no application to the real world. Once we have eliminated them we can put all that distressing

philosophical perplexity behind us. We can get on with more serious and workable pursuits, like natural science.

I assume this sketch of the philosophical lie of the land will not seem unfamiliar. Different philosophers at different times have found one or other position on the DIME shape attractive and settled on it. It is, I surmise, a common experience to find oneself moving from one location to another, as the demerits of that temporary resting-place make themselves felt. And there is a pattern to this dance – as D yields reluctantly to I, as I encourages a flirtation with M, as M propels one to E, as D seems like the place to try again. Plainly, too, E and D make natural partners, as do I and M. E is what opponents of D theorists accuse them of, and M is held to be the sub-text of the I adherent. It may not, in any particular case, be perfectly clear whether a given philosophical thesis is of the D or E variety, and similarly for its classification as I or M. E positions sometimes look like brazen versions of D positions, and M positions can seem like the logical conclusion of I positions. In any case, it should be clear enough for now how the philosophical choreography goes. It is a demanding piece, and it never seems to end.

We shall find this pattern repeatedly exemplified as we investigate the topics of the following chapters. And my thesis will be (a) that TN is a neglected alternative to any of the DIME positions and (b) that it is arguably preferable to those positions, especially for someone who already despairs (like me) of making good on a DIME-defined solution for philosophical problems. TN allows us to retain C-facts without underestimating or distorting them, without declaring them brutely inexplicable, and without courting ultimate mysteries in the world. We can thus escape E without being forced into any of the unsatisfactory positions to which E can seem the only way out. TN tells us how not to be eliminativists, while facing up to the deep and intractable problems C-facts pose to our modes of understanding. At the very least the TN option shows that it is a *non sequitur* to infer eliminativism from the failure of our epistemic faculties to comprehend what perplexes us. That would amount to the (idealist) fallacy of deriving an ontological conclusion from epistemological premisses. A better view, I shall suggest, is that the apparent

compulsoriness of the DIME dance results from systematically ignoring a TN position; so those who sense futility in that familiar sequence of steps are released by TN from having to participate therein.

IV The CALM Conjecture

Ideally, TN needs to be accompanied by a worked-out theory of human cognitive capacity, from which it would be demonstrable that certain forms of understanding are not humanly accessible, or run against the cognitive grain. This theory would be the analogue for the faculty of reason (whatever precisely that means) of a theory of the universal structure of human languages. That latter theory, when taken as a description of the human language faculty, contains principles that circumscribe the class of languages accessible to the growing child, since the mind incorporates a specialized language component that is selectively tuned to languages of the specified structure. Thus the grammar of human languages determines the scope and limits of the human language faculty, a particular organ of the mind.[11] What TN ideally requires, then, is something to play the role of grammar in delimiting what is accessible to reason, where this something fixes boundaries across which philosophical thought cannot travel. Needless to say, I have nothing to offer that is even remotely comparable to the present state of theorizing about the human language module; but it is important that some stab should be made at saying what at least such a theory would look like. So I shall now introduce a framework for thought, to be deployed in the following chapters and developed in chapter 8. The point now is just to get the basic idea across, so that we shall have something with which to test our intuitions in particular areas.

The CALM acronym stands for 'combinatorial atomism with lawlike mappings'. This is intended to capture a certain mode of thought, suited to certain subject-matters: that in which an array of primitive elements is subject to specified principles of combination which generate determinate relations between complexes of those elements. This combinatorial mode of thought, which yields a certain kind of novelty in the domain at issue, and proceeds in

bottom-up style, may represent contemporaneous relations between the structures dealt with, as well as dynamic relations over time. The essence of it is to yield understanding of the domain, especially its generative aspects, by means of transparent relations of composition between elements: we can see, on the basis of a CALM theory, exactly how – by what principles – items in the domain of study are related to each other. Put differently, if we already know, pretheoretically, that there exist principled relations between these items, a CALM theory tells us what the nature of these relations is – it specifies the manner in which the domain is structured. To grasp the theory is thus to understand the domain.

Now I have stated the CALM idea in an intentionally abstract way, not mentioning any specific subject-matter which conforms to it. However, it is not difficult to cite areas of theory that pretty clearly exemplify the pattern: physics, linguistics and mathematics have a CALM character. I shall be brief about why this is so. In physics we deal with elements laid out in space and subject to aggregative operations; the resulting complexes (macroscopic material objects) are then governed by lawlike relations which map successive states of the physical world onto one another. Physical 'novelty' is a function of the aggregative rules and the laws of change over time.[12] In linguistics, too, we conceive our domain in terms of primitive elements (words, phonemes) that come together to form complex wholes, where the properties of the whole are projectible from properties of the elements and their principles of combination. Speech consists in the production of these complexes over time, with determinate linguistic relations between them. The aggregative rules here are formulated in a grammar – the rules of syntax and semantics. Linguistic novelty is then explained in terms of the combinatory rules and the primitives they operate on; on the basis of our grammatical theory, we can see how to generate novel linguistic structures. In mathematics geometry provides the most obvious illustration of the CALM format; indeed, one might well think of the CALM structure in general as the geometrical mode of thought transferred to other domains. In geometry, clearly, one works with geometrical primitives – lines, planes, volumes – and combines these into ever more complex structures, with precise rules laid out as to how geometrical objects are related. Theorems are proved on the basis

of fundamental geometrical relations; the entire field has the kind of intelligible transparency we seek in understanding. But it is not just in the spatial (or quasi-spatial) parts of mathematics where CALM holds sway: number theory and set theory also fit the format. Elements and laws of combination prevail; mappings and functional relations abound; building up from the simple to the complex is ubiquitous. Spaces and structures and definable relations are what it is all about.

Of course, much more could be said in deepening and qualifying this abstract description of the fields mentioned, but I hope it is clear enough what the CALM idea is getting at. What will particularly concern me in what follows is a certain conjecture in which CALM features, namely that the philosophical problems we shall be dealing with resist resolution in CALM terms. We cannot dispel our perplexities by bringing to bear a CALM-style theory of the phenomena that puzzle us. The suggestion, then, tentatively made, will be that it is our conformity to CALM modes of thought that stands in the way of our achieving the kind of understanding we seek. That is the way our reason makes things intelligible to us, but in these cases the method breaks down, thus producing intractable puzzlement. In short, the CALM structure is to philosophical problems what human grammar is to nonhuman languages – an unavoidable but unsuitable mode of cognition. We apply the CALM mode willy-nilly to our problems, but instead of solving them it only steepens our sense of perplexity. That, at least, is to be the working conjecture.

V Philosophy and Common Sense

Commonsense knowledge can be divided into two parts: knowledge of matter and knowledge of mind. We know the basic properties of things in space and how to negotiate these things practically, and we also know the basic properties of psychological beings and how to negotiate them practically. Both kinds of knowledge are exceedingly primitive, not in the sense that they are intrinsically simple, but in the sense that they condition our thought and action from an early age, and are acquired spontaneously. Higher animals share this kind of knowledge, suited to their given environment, and human children are in possession

of it without benefit of explicit instruction. In many respects, the developmental characteristics of this knowledge mirror that of language: fragmentary data, rich system of knowledge, easy acquisition. It is thus highly plausible to suggest that commonsense knowledge has both an innate component and a modularized structure in the mind: the commonsense faculty has the kind of biological status attaching to the language faculty.[13] Nor is this surprising in view of what we know in general of the mind and given the evolutionary advantages of installing such knowledge in the organism's original endowment. Folk psychology, in particular, which will be our special concern, is plausibly viewed as a specialized subsystem of the mind, equipped with its own distinctive principles and programme of developmental expression, as well as a specific biological purpose. Let us, then, think of it as the mental equivalent of a physical organ of the body, with its own particular structure and function. Now we can ask how this cognitive organ might be related to the kind of reflective knowledge we seek in trying to do philosophy.

What is immediately striking, once one attends to it, is the enormous contrast that exists between the unreflective ease of acquisition of folk psychology, the ready manner in which we become adept in wielding it, and the extreme difficulty we experience in striving to make sense of, to explain, the basic ingredients of this knowledge-system. What came so easily to begin with cannot be made reflective sense of at maturity. It becomes as impenetrable as philosophy; in fact, it becomes philosophy. In the terms introduced earlier, we are programmed to employ concepts that are mysteries to us at a theoretical level. We can solve problems by *using* these concepts, but we cannot solve the problems they themselves raise – so says TN anyway. And this can seem surprising, for how can concepts that arose in us so smoothly be so resistant to reflective understanding?

The point I want to make is that this really should not seem so surprising once we have adopted the correct picture of the status of commonsense knowledge in the mind, and hence the truth of TN should not be seen as in any way paradoxical. Compare linguistic knowledge: would it be surprising if adult linguists proved unable to solve the theoretical problems posed by the output of the innate language faculty? No, because there

is no a priori reason why the component of mind that yields ordinary linguistic knowledge should be penetrable by the components of mind that seek reflective theoretical knowledge: what one mental organ can do, and do readily, may not be capturable or replicable by means of other mental organs. Similarly, if folk-psychological knowledge arises in a specific component of the mind, possibly dissociable from other components, then it is entirely possible that the component we use when trying to do philosophy should be unable to get very far in developing theoretical knowledge of the output of the folk psychology faculty. Different components of mind enter into different cognitive tasks – language, common sense, science, mathematics, philosophy – and there is no general expectation that the concepts available to one component will be transferable to another. It may thus be that folk-psychological concepts are inherently resistant to the kind of theoretical understanding proper to the reflective faculties of mind. In trying to understand folk psychology we are bringing one mental organ to bear on another, but this may be as futile as trying to pump the blood with the kidneys. Less colourfully put, on a modular view of human cognitive capacity, TN is by no means surprising, since it says merely that what comes easily to one faculty, for its limited purposes, may altogether defy the efforts of another faculty with *its* limited purposes. There is thus nothing very remarkable in the idea that we may not be able to understand the presuppositions of our own concepts. These presuppositions may simply not be the business of the common sense faculty, and they may not be accessible to the reflective faculties. It is not that somehow we *ought* to be able to answer the questions our commonsense concepts raise, if only we put in more effort.[14]

Still, if this is right, it does provide some account of the teasing, and even shaming, quality that philosophical puzzlement is sometimes felt to have. The problematic concepts enter our thought processes with great ease, but when we reflect on them we encounter deep difficulties. This can be felt as mildly embarrassing: we must be pretty dense to have so much trouble making articulate sense of what children pick up without a second thought. But actually, if the present picture is right, there is no reason for self-castigation: it is just that the contents of one module are not

explicable in the terms proper to another. It is not so very different from being unable to explain the physical workings of one's own body. So there is no real paradox in the idea that folk psychology is both remarkably easy and impossibly hard. Nor should it be simply assumed that philosophy can supply a theoretical vindication of common sense, if this means give an intelligible account of the place of commonsense facts in the world. Philosophy and common sense belong to different regions of the human mind, which may be related only tenuously.[15]

Let me make clear the scope of my aims in this book. I am not intending to discuss every question we dub 'philosophical'; I shall be dealing with a certain set of central questions, mainly relating to the mind. Whether the approach can be extended beyond this (already ambitious) set I shall not enquire. There is, of course, no necessity that every question discussed in departments of philosophy should be of the same underlying kind: intellectual natural kinds are not fixed by institutional demarcations. I do think, however, that the questions to be discussed in what follows do naturally belong together, so that a unified approach, metaphilosophically, is a sensible project. When I speak of 'philosophy', then, I should be understood as referring to this batch of questions, and possibly any others of the same type; it is of no concern to me that philosophers may in fact discuss questions for which TN is clearly inappropriate, or that nonphilosophers might discuss questions for which it is (say, physicists).

The topics that will occupy us in the ensuing chapters include consciousness, personhood, freedom, intentionality, knowledge. These notions are all embedded deep in folk psychology, so our general question is whether TN is the right perspective to take on such folk-psychological notions. Are the ultimate natures of the phenomena so signified open to our theoretical understanding? Are our own minds in principle intelligible to us?

NOTES

1 I am here following Noam Chomsky: see *Reflections on Language*, chapter 4, and *Language and Problems of Knowledge*, chapter 5.

2 This remains the case even if *any* mind (in some sufficiently well-defined sense) is incapable of answering the question at issue, since no ontological consequences follow from universal incapacity of this kind. It is still a point about *minds* – not the objective world – that they cannot, and cannot essentially, answer a certain type of question.

3 On mental organs and cognitive structure see Chomsky, *Rules and Representations*, chapter 1.

4 'Non-natural' is a catch-all term, encompassing a variety of philosophical trends, ranging from the explicitly God-invoking, to the 'queer processes' Wittgenstein speaks of, to certain kinds of ethical and mathematical realism. Not much weight should be placed on the intended general notion. TN is profoundly suspicious of the whole natural/non-natural contrast – except insofar as it characterizes the phenomenology of philosophical thought. What exists does so without impediment or metaphysical inharmony: it simply is.

5 For a defence of this kind of realism see Thomas Nagel, *The View From Nowhere*, chapter 6.

6 Again, this is a Chomskian thesis: see *Rules and Representations*, chapters 2 and 3. To be biased in favour of the universal grammar specific to human languages is, *eo ipso*, to be biased against grammars that diverge from this – richness in one direction going along with poverty in another (and contrariwise).

7 Of course, so-called *tabula rasa* conceptions of mind ultimately differ from structured conceptions only in degree, since no sense can be made of the idea of a wholly structureless cognitive system – any more than that of a formless physical object. Still, even if all minds were to share the same intrinsic nature, some problems might exceed the capacities of every such mind, thus generating cognitive closure without cognitive differentiation.

8 Remember, however, that such cognitive deficits are apt to be the inevitable outcome of cognitive strengths along other dimensions: we are bad at philosophy *because* we are good at something else – rather as we are bad at breathing under water because we are good at breathing in the open air.

9 Analytic philosophy, as a metaphilosophical position, is thus premissed on the assumption that the nature of certain objective facts is coded into the concepts we bring to those facts, so that philosophical truth is to be ascertained quite differently from other kinds of truth – as it were, by gazing into the conceptual mirror in which reality is reflected. This is actually, when you think about it, a very surprising and radical idea – by no means the platitude

its familiarity suggests. For why *should* certain parts of reality, and not others, be thus coded? Certainly this does not follow from the admission that *some* conceptual clarification is always part of philosophical enquiry. What is rather needed is the startling idea that some objective phenomena have already yielded up their inner nature to the human conceptual scheme – an idea that can hardly be regarded as axiomatic.

10 The distinction between science and philosophy, which is relatively recent in intellectual history, is, for TN, largely an artifact of the epistemic capacities we bring to bear on our problems: 'science' is simply the name we apply to questions that fit our theoretical faculties, while 'philosophy' denotes questions that do not. Strictly speaking, then, science never slices off what is properly called philosophy, i.e. that which has the kind of special hardness of which TN has an account. What happened historically was, in effect, that certain questions traditionally labelled 'philosophy' were seen to differ from others in point of their intellectual accessibility. So the history of thought can be seen as a kind of map of the human cognitive system, depicting its powers and limitations.

11 See Chomsky, *Rules and Representations*, chapter 6.

12 Of course, I am oversimplifying the content of extant physical theories, in all their exotic glory. My point is not that everything in physics is unmysterious because physics is comprehensively CALM; it is, rather, that physical concepts and theories are unmysterious *in proportion* as they have a CALM interpretation – and they fundamentally do. Quantum physics is theoretically problematic, at least in certain respects, just because it fails of CALM interpretation. Newton's original sense of the unacceptably occult character of the gravitational force, in contrast to other aspects of his theory, might be thought to have a similar source, since that force cannot be construed in terms of the rearrangement of constituent elements in a suitable medium. It is no part of TN, as I intend it, to assert that nothing outside what is commonly designated 'philosophy' presents problems of understanding comparable to those that typify philosophical questions of the kind I shall be discussing. So-called foundational issues in the sciences might well tap into the same biases and deficits that generate what we think of as philosophical perplexity. In general, the interesting distinctions here do not necessarily coincide with the usual institutional demarcations.

13 For a discussion see Jerry Fodor, *Psychosemantics*, epilogue.

14 Note that chimps possess commonsense psychological knowledge, useful in organizing their social relations, but lack the reflective

capacities we enjoy; so they cannot appreciate the problems raised by their own scheme of psychological concepts. They share one of our modules but not the other. Perhaps in the fullness of time they will evolve a reflective capacity, less developed than ours, and then become puzzled about their concepts; but we will not expect them to be able to lay their puzzlement to rest just because they are now equipped to feel it.

15 It is important here to distinguish between two vindicating projects: on the one hand, to protect common sense from philosophical perplexity; on the other, to answer such perplexity by producing a *theory* of common sense notions. TN says that the second project is impossible, but it refuses to infer that common sense should be abandoned, since it interprets the impossibility purely epistemically – so it serves in the less ambitious vindicating project. TN offers what I have elsewhere called a 'nonconstructive' vindication of common sense: see *The Problem of Consciousness*.

2

Consciousness

I The Problem: Consciousness and the Brain

Consciousness undoubtedly exists: it has the status of a datum, not a dispensable theoretical construct. But it ought not to be possible at all, given what we know of human and animal bodies, for there seems nothing about physical organisms from which it could conceivably arise. Physical states of the organism are, to all appearances, *de facto* necessary and sufficient conditions for conscious states, and the brain is surely centrally implicated, but electrochemical impulses travelling along nerve strands seem far removed from what they somehow secure. The operations of matter look like a singularly inadequate foundation for a mental life – a plan for making conscious states that stands no chance of success. It is thus numbingly difficult to make sense of the fact of material emergence, since nothing plausible suggests itself as an adequate basis for getting consciousness off the ground. Intelligibility expires in the explanatory vacuum that confronts us. We have something like physical supervenience, but this only accentuates the explanatory problem rather than solving it, since the supervenience appears brute and unmediated. We have no conception of what a unifying theory of consciousness and matter would look like. The resulting logical gulf presents us with a deep mystery: how does the world contrive to do what we cannot conceive of it as doing? That is the mind–body problem – finding an explanatory theory of the psychophysical link that will enable us to resolve the mysteries the data present.

This is a simple and pure form of the kind of problem characteristic of philosophical perplexity. Common sense (perhaps augmented with a bit of science) tells us that something is the case, but we have the greatest difficulty in developing a conception of the world that will allow us to accept what common sense tells us. It can then appear that we shall be compelled to revise common sense, on pain of absurdity in our conceptual system, or metaphysical conjuring tricks in objective reality. Consciousness puzzles us in a special way – the way we label 'philosophical': it is quite unlike our puzzlement over (say) how salt dissolves in water or plants grow.[1] The head spins in theoretical disarray; no explanatory model suggests itself; bizarre ontologies loom. There is a feeling of intense confusion, but no clear idea about where the confusion lies. It is also a puzzlement that is easier to experience than to formulate, since it is exceedingly difficult to say precisely *what it is* about consciousness that makes it so uncongenial to physical explanation. Our intuitions outrun our diagnostic powers – in a way that is also characteristic of philosophical bewilderment. Something is wrong somewhere, deeply so, but even putting one's finger on it can prove testing. Hence there exist philosophers who deny (with eyes studiously averted) that there is any real problem about consciousness.

In this chapter I shall apply the apparatus introduced in the previous chapter to the problem of consciousness. This should serve to make that apparatus more concrete, as well as to shed light on the particular problem at hand. My further aim is to treat that problem as a philosophical paradigm, approaching the later problems with this as a model and guide. First we shall see how TN applies to the mind–body problem; then we shall extend it into neighbouring areas of philosophical aporia.

Let me begin, though, with matters of formulation, so that we are as clear as we can be about what we are asking; a little preliminary pedantry may ease some of the murkiness of the topic. What, then, is the mark of a conscious state? Where precisely is the problem located? A celebrated stab at encapsulating the property of consciousness that eludes physical explanation introduces the phrase, 'what it is like to be a K'.[2] This is intended to capture the intrinsic non-relational essence of the conscious state, the aspect that distinguishes it from states that may share

a similar set of extrinsic relational properties. To satisfy such a description is to enjoy subjective states, states which exist *for* a subject. Now this is certainly an apt locution, but it can mislead, so let me warn against some possible misconstruals of it.

First, let us attend to questions of logical scope. Consider the sentence, 'there is something it is like to be a bat': this contains two quantifier expressions, at the beginning and end. The first, 'there is something', is best taken as second-order, ranging over properties that bats may instantiate, so that the entire sentence can be cumbrously parsed as 'there is some property P such that bats have P and P confers "likeness" on bats'. P is a subjective property of bats, in contrast to bat properties that confer no 'likeness' on them. In any case, we are quantifying over a type not a token, a universal not a particular: no conscious token confers any 'likeness' beyond that conferred by its conscious type. The second quantifier phrase, 'a bat', is plainly first-order and ranges, universally, over individual bats. We have to be careful about the relative scopes of the two quantifiers if we are to read the vernacular phrase in the intended way. The claim is not that, concerning any individual bat x, there is something it is like to be x; the claim is rather that there is something it is like to be any bat at all. The subjective property is not tied to the particular bat we happen to be considering but encompasses all (normal) bats. It is not that being a *particular* conscious subject has its own unique subjective type, so that mine might be different from yours simply in virtue of our numerical distinctness; rather, my subjectivity is shared by any being relevantly similar to me. Phenomenologies are not individuated by numerical identity; they concern qualitative identity. So there is no suggestion here that the problem of subjectivity has to do with distinguishing one individual subject from another (though this latter is a real problem in its own right, of course). It is about the experience-type common to the generality of bats (say).[3]

That was a fairly straightforward point; the next one cannot be sorted out quite so crisply. It concerns the notion of what a certain experience-type is like *for* its subject. It is tempting to read the phrase as suggesting that there is a way that *being* a bat is for a bat – a way bat experience strikes bats. On this understanding, we are saying that the experiences are presented to the

bats in a certain way, that bats take their experiences as objects of apprehension; and thus the subjective character of experience is held to consist in how experiences appear to something like introspection. And if that is so, then the bats must somehow represent their own experiences to themselves, bringing them under higher-order intentional states. They must possess states whose content refers to the experiences delivered by their sonar sense; and it is the possession of these states that constitutes subjectivity.

Clearly, something is going wrong here. We don't want to credit bats with higher-order representational states, i.e. self-reflection; or if we do, it is not merely because we think their sensations have subjective character. For surely sensations have a subjective aspect whether or not their subjects can reflect on them and acknowledge this fact. What is presented to the bat in a specific subjective way, when its sonar sense operates, are things in the external world, the environmental objects it perceives by means of that sense. What it is like to *be* a bat is identical with what the *world* is like for a bat. The bat's subjectivity consists in the particular way in which the perceived environment appears to the bat, not in how those perceptions themselves appear to it. The only intentional contents here represent external objects, but the manner of this representation confers a subjective character on the perceptual experiences that bear this outer-directed content. Thus in the specification of the bat's subjectivity the only intentional relation involved holds between the bat and the world, not between the bat and its own experience *of* the world. It is a matter of how those rebounding high-pitched sounds appear to the bat. In other terms, it is a matter of the secondary qualities associated with the bat's sonar sense.[4]

Then we can formulate our perplexity about consciousness as follows: how is it possible for states which there is something it is like to have to arise out of states of a kind which there is nothing it is like to have? The physical states that correlate with conscious states – neural firings of certain frequencies – look, on their surface, to be states of a kind that can occur in the total absence of consciousness; indeed, there seems nothing about them that could explain why *those* states, rather than (say) states of the kidneys, are the basis of consciousness. One could never tell, just by inspecting brain states, that they are uniquely the source of

consciousness; this is something we know only by independently established correlations. It is as if *any* kind of physical state could have turned out to be the basis of consciousness; it just happens that it is neurons and their peculiar activities. Thus there is an irreducible bruteness to the correlation, as if consciousness has just been *pasted on* to the cerebral material. By some unknown process, electrochemical events give rise to states which there is something it is like to have: a subject of awareness is bodied forth from raw materials that look remarkably unsuitable for the job (not that we have any idea what other sorts of materials would be cut out for the job). The problem is essentially architectural: how would you set about constructing subjective states from the cellular structures that compose the brain? Until we have some idea how to answer that, and in particular some grasp of the architectural principles involved, as we do for other biological traits and organs, we are faced with a gaping explanatory hole in our theory of how the world works. That hole is called 'the mind–body problem'.[5]

II DIME and Consciousness

My programme, I said, is to illustrate the DIME alternatives using consciousness as a philosophical paradigm, and then proceed on this basis to other topics. I shall not be attempting to show, in this chapter or later ones, that the range of DIME positions philosophers have entertained is inadequate to the problems; this is indeed my opinion, but I do not expect to establish it here. Brief surveys must suffice, accompanied by tendentious indications of where the weaknesses lie. Adherents of these positions cannot expect to be persuaded out of their convictions. I speak, rather, to those who are similarly dissatisfied with the usual sorts of positions and would like to see a way out. There is, I believe, a systematic pattern to philosophical disputation, in which the same kinds of unsatisfying alternatives recur; the point of TN is to break the hold of this pattern, by supplying a better alternative and by making metaphilosophical sense of the usual dialectic. The tone, then, will be diagnostic rather than refutative. So: into what kinds of (contorted!) posture has consciousness driven philosophers, past and present?

Domesticating programmes are familiar enough – attempts to convince us that consciousness is really nothing more than such-and-such. When you analyse conscious states sufficiently the specialness dissolves. Consciousness can be reduced to facts of a metaphysically unproblematic kind. Materialism and functionalism are the most obvious D positions today: to be in a conscious state is just to be in a certain sort of physical state – a neural state or a state defined by causal role.[6] The spookiness is an illusion, to be dispelled by acquiring more physical knowledge of the kind we already possess. Our conceptual scheme already contains the essential resources for a comprehensive theory of consciousness. Also to be included under this heading are such ideas as that consciousness is just a kind of self-monitoring or higher-order belief state or criterionless self-ascription;[7] or again, that it is simply one kind of emergent biological property among others, raising no deeper question than that raised by the nature of digestion or the like.[8] D positions thus take something relatively commonplace and well-understood, something less mysterious-seeming, and assert that consciousness can be explained in those terms. The solution to the mind–body problem is then consequent on this assimilation.

The standard and oft-repeated objection to such positions is simply that they fail to do justice to the facts: it is just not plausible that consciousness is nothing more than the things that are thus held to constitute it. The reductions miss something out, the essence indeed. This general intuition is then often backed up with specific objections, which typically have less power than the general sense of inadequacy: that the psychophysical link is more contingent than these theories allow;[9] that specific aspects of phenomenology elude explanations of these kinds;[10] that we can imagine creatures who satisfy the reductive conditions yet lack any consciousness at all.[11] In short, domestication is tantamount to denial, to defying the data.

Irreducibility theses suggest themselves when once the prospects for domestication dim. If consciousness cannot be reduced to something familiar, that must be because it is intrinsically irreducible – ontologically basic, an explanatory terminus. Accordingly, we must accept that psychophysical correlations, biological emergence and physical supervenience are all simply brute facts, admitting of no explanation. The nature of consciousness is already fully represented in our ordinary concepts, and it is only an exaggerated explanatory urge that makes us think our theories of the world have any essential incompleteness. It is simply an inexplicable fact that irreducible conscious states have the kinds of relations to the physical world that they have. There is nowhere deeper to dig; the world has no ontological complexity beyond that recorded in our ordinary descriptions of things. Mental properties have no internal intelligible relation to physical properties, despite their dependence on such properties. We must accept the duality without perplexing ourselves about its possibility.[12]

Here the objection is apt to be that brute irreducibility, while paying ample respect to the *sui generis* character of conscious states, does so only at the cost of rendering their place in the world unacceptably mysterious; it thus abnegates our explanatory responsibilities. What is it about neural tissue in particular that makes *it* capable of subserving conscious states? Would we be content with the claim that consciousness emerges from sawdust quite inexplicably? Are we not mistaking human ignorance for ontological basicness? The leap from matter to mind is surely too great to be totally unmediated; it must be backed by natural principles of some kind. And if, as seems plausible, there are objective necessities at work in tying consciousness to the physical world, there must be some account of these necessities; it isn't merely accidental or adventitious that the brain is the organ of consciousness. Coincidences may not need explanation, but it is surely no coincidence that brain tissue and conscious processes go systematically together. Irreducibility theses are culpably silent on such questions. They leave consciousness hanging.[13]

Miracle theses have tended to lapse in these secular times, at least in scientific circles, but they were more or less orthodox until relatively recently. They may be divided into theses in which a divine being is brought in to underwrite the miracle and theses in which the miracle is taken as ultimate. Of course, it is quite unclear that this notion of objective miracle is even coherent, but that is not sufficient to deter people from subscribing to ideas framed in these terms. The thought (or attempted thought) is that the world is not fully intelligible in terms of causes, laws, mechanisms, natural forces; there are ultimate anomalies out there,

contraventions of the naturalistic viewpoint. God plays tricks with nature, or nature plays tricks with itself. Thus we have the traditional idea of the soul, an immaterial particle (sic) floating somewhere above the flux of physical events. It owes its being to nothing (except perhaps God), it is immortal, and it can recur in subsequent lives. Its interaction with the body is miraculous, quite possibly requiring God's continual intervention. It belongs to that order of reality in which angels and ghosts and miraculous healings occur. It is not of this world. You get goosepimples just thinking about it! It is futile to try to explain it in sublunary terms; its nature is to flout nature. It is, precisely, *super*natural. Consciousness is the divine spark in each of us.

I take M positions seriously, not as genuine candidates for truth, but as expressive of the philosophical hysteria that so readily envelops us. What is interesting is that we can find ourselves uttering these words, or falling inchoately into these thoughts. For it is doubtful that any of this really means anything. It is mere poetry, rhetoric, word-spinning. Falsity is not the main problem, though doubtless there is some of that; the problem rather is that of coherence, of staking out a genuine position in logical space. What could it *mean* to say that consciousness is supernatural? What content does the notion of the supernatural really have? Still, radically defective as M positions no doubt are, they exercise a powerful hold on speculative reason, so we must include them among the responses that philosophical problems provoke. They play their part in defining the options to which thinkers resort in contemplating the mind–body problem (among other things). And showing how to avoid adopting an incoherent position is often a substantive philosophical task. Indeed, one of the chief merits of TN is that it allows us to dismiss all such intimations as by-products of our principled ignorance on the matter, faltering steps into the epistemic abyss.[14]

Eliminative theses may be regarded as positions of last resort. When D ambitions have been abandoned, and I declarations have come to sound hollow, and M creeds have been forsworn, *then* it starts to seem compulsory to reject the thing that generates the problem. Boldy one asserts: there is *no such thing* as consciousness. One undertakes to eliminate it from one's ontology; one encases talk of it in inverted commas. By so denying its very existence one sidesteps the mind–body problem altogether: that problem is the pseudo-problem of trying to link a mythical realm to sober reality. If there were such a thing as consciousness, it would have to be magical; but there is no magic in the world, so neither is there consciousnes. Less extremely put, there is no room for consciousness in our emerging scientific view of the world; and what resists scientific integration had better be eliminated altogether.[15]

The usual response to eliminative theses is plain incredulity: to deny that one is concious requires one to deny what is self-evident. It is not like denying the existence of vital spirits or the devil, since conscious states are *data* – part of what the world presents to us as simply so. Moreover, to cease to talk in terms of consciousness would be to cripple our entire conception of ourselves and one another. E theses are in the position of rejecting the obvious because no good theory of it can be found.

Thus it is that the DIME shape stamps itself onto the topic of consciousness. Now we shall ask how TN responds to the problem.

III TN and Consciousness

TN with respect to consciousness is this claim: the natural principles which mediate between brain processes and conscious states are inaccessible to human reason. We would need a conceptual revolution in order to solve the mind–body problem, but it is not a revolution our intellects can effect. It is a general trait of organisms to have areas of cognitive strength and weakness, and the human cognitive system is weak precisely where the problem of consciousness arises. The requisite theory does not come within the scope of our mental modules.

According to TN, the DIME shape characterizes philosophical debate about the mind–body problem precisely because of the truth of TN. D projects tempt us, and predictably fail, because we try to force conscious phenomena into a conceptual mould that ill suits them, striving to bring them under a set of theoretical notions that is available to us – but actually doesn't fit the facts. The correct theory lies to the side of what we can generate,

so we make do with theories that at best approximate to the truth. I positions attract us because no accessible theory offers any explanatory hope, so we rush to deny that any such theory exists – thus fallaciously deducing an ontological conclusion from premisses about human epistemology. M doctrines are hyperbolic responses to the (epistemically) mysterious character of consciousness: they are reifications of our own cognitive limitations. And E conclusions are panicky attempts to remedy what is ultimately an epistemological problem: if we cannot understand it, even in principle, then we are prone to deny that it exists. In that way we can protect ourselves from the unflattering truth that parts of nature will not yield their secrets to the human cognitive apparatus.[16] The organ sitting in our heads has not the size and power to comprehend everything that exists.

Thus TN predicts that the DIME shape will imprint itself on the mind–body problem, but it denies that DIME exhausts our options. And in so far as those options are admitted to be unsatisfactory, TN offers itself as a preferable alternative. It is, at the very least, a hypothesis worth taking seriously; a factual hypothesis, indeed, about the structure and scope of human understanding, to be evaluated as empirical hypotheses generally are. What, we should ask, is its antecedent likelihood, given the general nature of evolved cognitive systems; and what evidence from the field of human enquiry might speak in its favour?

Now I have defended TN about the mind-brain link at some length in *The Problem of Consciousness*,[17] and I do not propose to repeat here everything I said there. My aims now are more illustrative: I want to use the mind–body problem to exemplify the general metaphilosophical position I am exploring, and to provide a paradigm for other philosophical problems. Let me then quote a remarkable passage from the nineteenth-century scientist John Tyndall, which succinctly expresses the spirit of the TN position: 'The passage from the physics of the brain to the corresponding facts of consciousness is unthinkable. Granted that a definite thought and a definite molecular action in the brain occur simultaneously, we do not possess the intellectual organ, nor apparently any rudiment of the organ, which would enable us to pass, by a process of reasoning, from one to the other'.[18] This gets it exactly right by my lights, even down to the suggestion of

an explanation of the unthinkability in terms of mental modularity. My general thesis, in these terms, is that philosophical bafflement results from the lack of an 'intellectual organ' suitable to the subject. I would only add, what Tyndall leaves implicit, that the unthinkability is no reason to suppose that anything supernatural or intrinsically brute is going on: this is strictly a point about how our intellects are constituted, not a comment on the miraculous doings of the real world. That *we* cannot make the explanatory passage from brain to mind does not entail that the *brain* exercises any magical function in so doing. Epistemic limits never entail ontological fissures or fishiness.

Can we give more colour to this idea of organ lack? What is it about our modes of thought and our access to the phenomena that generates the closure TN detects? Here we can do little more than point to clues, speculatively interpreting what we find. I shall mention two points, both suggestive rather than apodictic; this should at least tell us the *kind* of thing we should be seeking in diagnosing the truth of TN. The first point concerns CALM, the second a particular property of introspection.

It is a familiar thought that conscious states resist emergent explanation in terms of mereological notions: that is, we cannot think of pains (say) as aggregates of the neural units that underlie them, either cells or the firings thereof. By contrast, higher-level properties of liquids (say) can be construed in terms of lower-level constituents and their combinatorial possibilities: so we understand the relation (supervenience, in effect) between the higher-level properties and the underlying matrix of combining molecules. This is, in my terms, a pure case of CALM understanding: atomic elements combining according to certain laws and mapping intelligibly onto the facts to be explained – parts and wholes, basically. But this is just what we are prevented from doing in the case of consciousness and the brain: conscious states are not CALM-construable products of brain components. Here the mappings, which must exist in some form, are inscrutable in CALM terms. We can readily conceive of higher-level brain functions in terms of simpler composing constituents; but once we think in terms of consciousness this mode of explanation lapses. Thus we have no model of what the emergence relation might consist in; here the supervenience is opaque and puzzling, not

transparent and intelligible. Similarly, it seems quite unpromising to adopt a more syntactic CALM explanation: even if there are symbols in the brain, conscious states are not explicable as mere syntactic strings of such symbols. If conscious states have something like constituent structure, that lies at the conscious level itself; it is not a way of levering consciousness out of brain properties. Given that the CALM format governs our conception of natural emergence, it is no surprise that consciousness should be so baffling to us. We have a CALM bias, but we cannot implement this in explaining the mind–brain link. TN takes this to be symptomatic of the closure it alleges.

The second point concerns what I have elsewhere called the hidden structure of consciousness.[19] The basic idea is that conscious states conceal a hidden nature that enables them to hook onto brain states. Now this notion of the hidden is to be interpreted purely epistemically, as a point about the faculties we bring to bear in apprehending conscious states; it is not meant to be some kind of objective occlusion, as with tree roots buried under ground. So we can ask what properties of our consciousness-apprehending faculties might generate this kind of partial access. And there is a feature that is at least suggestive, which I shall call the 'single-channel' property of introspection. Let us agree that the scope of a cognitive faculty is constrained by its pattern of causal sensitivities: what it can represent is a function of the properties it can causally resonate to. This implies that the more fixed and invariant the causal dependence between states of the faculty and states of its (intentional) objects the less the faculty will tell us about those objects, other things being equal. If a faculty is operationally tied to a single perspective on an object, then it is unlikely to yield the whole nature of that on which it has this fixed perspective. My hypothesis, then, is that introspection is a highly restricted and rigid epistemic resonator; it is a single-channel faculty, confined to a mere subset of the properties of its objects. We do not enjoy a rich variety of modes of apprehension of conscious states, analogous to the five senses we bring to the external world, and the single mode we do have is notably inflexible in its operation. Compare vision or touch, which provide multiple causal channels onto their objects, corresponding to shifts of position and focus and so on. If we want to know about the properties of a chair, say, we can explore it from many points of view, using sight and touch, revealing new aspects as we proceed. Thus we develop a rich conception of its nature, and science becomes possible. But introspection does not similarly provide for a rich conception of its objects. If we want to know about a pain, there is little we can do but detect its presence by simple introspection. We cannot shift viewpoint or bring to bear another sense. We quickly run out of things to say about our conscious states because introspection tells us so few things about them. It is a bit like trying to discover the full nature of a chair with only vision and one's eyes permanently fixed two inches away from the surface of one of the arms! At any rate, that is the hypothesis we are entertaining. The thought then is that we should not be surprised that consciousness has a hidden nature, postulated for theoretical reasons, once we notice that our faculties of self-knowledge are trained upon it in such a restricted and uniform way. Introspection is remarkably effective in employing its single mode of access to detect *some* properties of conscious states, but this very fact makes it inept at developing a rounded picture of the objects on which it reports. Its operational principles do not suit it for revealing all the interesting properties of conscious states. Again, TN sees in this a (partial) rationale for the deep ignorance that afflicts our understanding of the phenomena. We can begin to see why the closure conjectured by TN should not be exclaimed over – it is a natural upshot of constitutive facts about the faculties that are relevant to the problem. Closure is not, of course, *proven* by these facts, but they do serve to make sense out of an acknowledged futility.

There is, after all, no a priori reason to suppose that the nature of consciousness is fully revealed to conscious beings themselves. On the contrary, consciousness was presumably designed chiefly as a vehicle of mental representation, not as an object of it: its job is primarily to act as a medium of thought and perception, specifically in respect of the external world. But when we try to form representations *of* consciousness, making the vehicle into its own object, we encounter a notable paucity: our concepts of consciousness do not lead the way into a developed science of consciousness. It is, as it were, a good object-language but an indifferent meta-language, lacking the resources to describe itself

with any degree of depth. Just as a natural language could express extensive knowledge of the world without having much to say about itself, so consciousness, as a medium of intentionality, seems pretty powerful about what lies outside it but provides little or nothing in the way of real theory about its own nature. This is obvious enough for the simpler conscious organisms, whose reflexive representations are minimal indeed, and there is no reason to suppose that human beings are in a qualitatively different case. As ordinary speakers lack rational mastery of linguistic theory, so conscious subjects lack a theory of consciousness; and we should not be terribly surprised if the lack is permanent, in view of the structure and function of the system.

IV Sense, Reference and the Mind–Body Problem

Using the sense/reference distinction, TN about the mind–body problem can be stated thus: brain states and conscious states fall under senses such that (i) under those senses the link between them is intelligible and (ii) those senses are not potential constituents of human thoughts. That is, the references of mental and cerebral terms have aspects, corresponding to (ideal) senses, which provide the kind of natural nexus we cannot, under our present concepts, envisage. We might say that these senses are the mystery-resolving senses for the philosophical mind–body problem; they correspond to the kind of conceptual shift that would render the psychophysical relation transparent, *if* it could be achieved. They occur in the propositions that constitute the (ideal) scientific theory of mind and body. And they are not identical with the senses we now associate with our terms; indeed, they must be far removed from these senses – though presumably some explanatory link has to hold between the two. If we imagine creatures whose cognitive structure allows them to incorporate the mystery-resolving senses into their thoughts, then we can say that for these creatures there is no aura of impenetrable mystery surrounding the psychophysical link. For them, the connexion is as unmysterious as any other natural nexus, a matter of plain science. TN thus diagnoses the character of the philosophical problem for us as consisting in the cognitive inaccessibility of the right senses, the ones that convert the problem into regular science.

On this conception, then, no *reference* ever in itself poses a philosophical problem: the objective world is philosophically unproblematic. Philosophical problems arise from the senses under which we conceive the world; they are, in one good sense, purely conceptual problems. By varying the senses we can transform a question from philosophy to non-philosophy, as when we imagine creatures whose modes of conception present them with no philosophical mystery; and we can do the converse too. Whether a question counts as philosophical depends upon who is doing the philosophizing, i.e. what conceptual and theoretical resources they possess: that is the point of TN. In the case of the mind–body problem, the mystery exists only for creatures whose cognitive slant biases them away from the concepts that are needed to make the question into a mere scientific problem. The mystery does not attach to the reference of mental terms *no matter* how this reference might be presented to a creature.

The point of my restating the TN position in these terms is to raise the following issue: is there such a thing as a *philosophical* answer to a philosophical question? And, given what has just been said, the reply would seem to be in the negative. For an answer to a philosophical question is a proposition, referring to the entities originally puzzled over, which contains senses that generate no peculiarly philosophical perplexity. Let us call this a 'scientific' proposition: then we can say that the answer to the mind–body problem consists in a set of scientific propositions, not humanly accessible according to TN, that are such that *were* they to be grasped by some being they would produce no sense of philosophical mystery in that being. They would have the same kind of epistemic status that the propositions of human science have. So the philosophical mind–body problem does not have a peculiarly *philosophical* answer; the theory that resolves the problem is not a distinctively philosophical theory. And in so far as this problem is a paradigm for other problems, they too lack distinctively philosophical solutions. If a class of creatures found digestion philosophically problematic, by dint of a conceptual lack that we simply take for granted, then the answer to their puzzlement would be a straight scientific theory of digestion, not some peculiarly philosophical theory of the nature of digestion. That is how TN views the human epistemic predicament with respect to

our (so-called) philosophical problems. We have philosophical questions, individuated by the senses we bring to our terms, but these questions do not have philosophical answers – except in so far as they answer philosophical questions. Knowledge of the world, including the parts of it that produce philosophical puzzlement in human beings, is all of a piece; where obtainable it all forms a continuous fabric of understanding. There is not philosophical knowledge on the one hand and scientific knowledge on the other. It only seems that this is so because of the deep partition among questions that results from our cognitive biases. In a sense, then, TN does not believe in philosophy as a separate discipline. There is, to be sure, such a thing as philosophical ignorance, but there is no such thing as philosophical knowledge – not as traditionally conceived. Put less grandly, the mind–body problem (for example) has a merely scientific solution, but it seems to us like a peculiarly philosophical problem because the requisite theory lies outside of our cognitive bounds. Thinking of the matter in terms of sense and reference can help to clarify what is going on here. In a slogan: philosophical problems attach to the level of sense not to the level of reference.

I doubt that many readers will feel persuaded of this claim at the present stage of the discussion, either generally or for the specific case of consciousness; but it should at least be coming clear what the claim is and what motivates it. My persuasive intentions are cumulative: assent at any one point will depend upon the appeal of the larger picture, upon the overall pattern that emerges from a number of areas. Does the TN hypothesis make things fall generally into place? For we are dealing here with a family of problems, interlinked in various ways, and what we say about any one of them will have a bearing on the right view of the others. We need a global perspective if TN is to be properly evaluated in more local contexts.

Two big links should be borne in mind from now on. First, TN with respect to the mind–body problem may be bolstered by considering connected areas in which a similar view appears plausible. I suspect that many people believe that the mind–body problem is in principle soluble in terms of future human science, and that I exaggerate our cognitive blankness on the question. Such people may, however, be less inclined to this kind of optimism (or lack

of pessimism) with respect to certain other philosophical problems I shall discuss, say the self or free will or meaning; so considering consciousness in conjunction with these may instill some doubts about its accessibility to future scientific understanding. For the question will need to be addressed as to why the consciousness problem is more tractable than those others, especially in the light of the affinities between them.

Secondly, consciousness is generally presupposed by the other mental phenomena we shall be discussing. So the problem of consciousness infects these other problems, spreading its intractability upon them, and giving them their special character. Those philosophers, then, who sense a scintilla of truth in my approach to consciousness (not those of the previous paragraph) may find their sympathies widening into other domains. The mind–body problem will be seen as not just a paradigm but a pervasive presence. In any case, it is the whole family of problems that needs to be treated, not each member in isolation.

NOTES

1 Perhaps we need a further distinction, which I shall mention but not pursue here – that between a puzzle and a mystery. It appears that we sometimes possess adequate theories that contain puzzling elements, as (say) with Newtonian theory and gravitation; this type of case is unlike that in which the domain of interest is deeply resistant to adequate theorizing, as (say) with consciousness. In the former type of case our cognitive capacities permit us to formulate the relevant theories but we cannot expel all sense of puzzlement about how the world is working; while in the latter type of case we cannot so much as formulate an explanatory theory that might contain puzzling pockets. TN can, in principle, be defined so as to admit both sorts of epistemic occlusion – localized puzzles and thoroughgoing mystery – but we shall be concerned with the second of these in this book (not that the distinction is obvious in every case). (I am indebted here to some suggestions of Chomsky made in correspondence.)
2 See Brian Farrell, 'Experience', and Thomas Nagel, 'What is it Like to be a Bat?'.
3 I have always taken this to be obvious, but confusion on the point is perhaps sufficiently prevalent to warrant some spelling out.

4 Similarly, the subjectivity of the human visual sense consists in the secondary qualities associated with that sense – *viz.* colours – and not in the way visual experiences strike our faculty of introspection, since the experiences would still possess subjective character even if we had no such higher-order faculty. The subjectivity of perceptual experience is a matter of how the *world* is perceived. See my *The Subjective View* for more on this.

5 It might be asked why consciousness is more problematic than magnetism or gravitation – these also being properties of matter that are puzzling to common sense. Part of the answer is that consciousness is an emergent or supervenient phenomenon, so that some bottom-up account ought to be possible of how it is produced; another part of the answer is that conscious states play no explanatory role in our developed theories of the material world, unlike magnetism and gravity. No doubt matter has its puzzling properties, but consciousness introduces a whole new dimension of puzzlement to that which exists independently of it. That is why dualism is a natural response to the fact of consciousness, as it is not for other properties of matter that may perplex us.

6 For a representative collection of essays, see Ned Block, *Readings in the Philosophy of Psychology*.

7 On consciousness as second-order belief, see David Rosenthal, 'Two Concepts of Consciousness'. The idea of criterionless self-ascription as definitive of a conscious state is a kind of Wittgensteinian version of essentially the same conception.

8 See John Searle, *The Rediscovery of the Mind*.

9 See Saul Kripke, *Naming and Necessity*.

10 See Wilfred Sellars, 'Empiricism and the Philosophy of Mind'.

11 See Ned Block, 'Troubles with Functionalism'.

12 See Donald Davidson, 'Mental Events', for a version of this approach.

13 If a straight irreducibility thesis were true in this case, then there should be no felt mind–body problem: the brain's possession of conscious properties ought to be no more puzzling than the fact that physical objects can have both shape and weight, each of these properties being irreducible to the other. But the existence and salience of the mind–body problem is a measure of the inadequacy of this kind of model. How *could* the brain have both material and mental properties? And in virtue of what do the latter properties depend upon the former?

14 Aphorism: 'the supernatural is human ignorance reified'.

15 Thus we had eliminative behaviourism at the beginning of the twentieth century and we have eliminative materialism toward its end: see Paul Churchland, 'Eliminative Materialism and Propositional Attitudes'. Like most brands of optimism eliminativism takes the form of denial.

16 But remember that incapacity in one domain is the natural upshot of fluency in others: see Chomsky, *Language and Problems of Knowledge*, chapter 5. A cut diamond will fail to reflect light in certain directions – but only because it glitters impressively in others.

17 I should note that in that work I was more inclined to view the mind–body problem as uniquely subject to TN; subsequent reflection, as reported in the present work, has convinced me that TN has broader philosophical application than I at first thought.

18 Quoted in William James, *Principles of Psychology*, volume 1, p. 147.

19 See my, 'The Hidden Structure of Consciousness'.