

PHILOSOPHICAL LECTURE

CONSCIOUSNESS AND CONTENT

By COLIN McGINN

*Corpus Christi College, Oxford*

*Read 29 November 1988*

NATURALISM in the philosophy of mind is the thesis that every property of mind can be explained in broadly physical terms.<sup>1</sup> Nothing mental is physically mysterious. There are two main problems confronting a naturalistically inclined philosopher of mind. There is, first, the problem of explaining consciousness in broadly physical terms: in virtue of what does a physical organism come to have conscious states? And, second, there is the problem of explaining representational content—intentionality—in broadly physical terms: in virtue of what does a physical organism come to be intentionally directed towards the world? We want to know how consciousness depends upon the physical world; and we want to know, in natural physical terms, how it is that thoughts and experiences get to be *about* states of affairs. We want a naturalistic account of subjectivity and mental representation.<sup>2</sup> Only then will the naturalist happily accept that there are such things as consciousness and content.

<sup>1</sup> This is the standard contemporary view of naturalism. See, e.g., Jerry Fodor, *Psychosemantics* (The MIT Press, Cambridge, Mass., 1987), Chap. 4. I do not say that it is my view of what it takes to be a good naturalist. As will become clear, I think we can view the mind naturalistically without being able to offer broadly physical explanations of its powers. (I say ‘broadly physical’ in order to include biological properties and higher-order causal properties, as well as the properties directly treated in physics.) An alternative way of putting the naturalistic demand is this: explain *why* it is that the mental is supervenient on the physical, given that it is. The general motive behind such naturalism is the avoidance of some sort of radical ‘emergence’ of the mental with respect to the physical. See Thomas Nagel, ‘Panpsychism’, *Mortal Questions* (Cambridge University Press, 1979), on why emergence is to be avoided.

<sup>2</sup> A third, and connected, problem is explaining how a physical organism can be subject to the norms of rationality. How, for example, does *modus ponens* get its grip on the causal transitions between mental states? This question is clearly connected with the question about intentionality, since rationality (as

Recent years have witnessed a curious asymmetry of attitude with respect to these two problems. While there has been much optimism about the prospects of success in accounting for intentionality, pessimism about explaining consciousness has deepened progressively. We can, it is felt, explain what makes a mental state have the content it has; at least there is no huge barrier of principle in the way of our doing so. But, it is commonly conceded, we have no remotely plausible account of what makes a mental state have the phenomenological character it has; we don't even know where to start. Books and articles appear apace offering to tell us exactly what mental aboutness consists in, while heads continue to be shaken over the nature of consciousness. Indeed, standard approaches to content tend simply to ignore the problem of consciousness, defeatedly postponing it till the next century. True, there are those rugged souls who purport to see no difficulty of principle about consciousness; but among those who do appreciate the difficulty there coexists much optimism about content. This is curious because of the apparently intimate connexion between consciousness and content: intentionality is a property precisely of conscious states, and arguably only of conscious states (at least originally). Moreover, the content of an experience (say) and its subjective features are, on the face of it, inseparable from each other. How then can we pretend that the two problems can be pursued quite independently? In particular, how can we prevent justified pessimism about consciousness spreading to the problem of content? If we cannot say, in physical terms, what makes it the case that an experience is *like* something for its possessor, then how can we hope to say, in such terms, what makes it the case that the experience is *of* something in the world—since what the experience is like and what it is of are not, *prima facie*, independent properties of the experience? That is the question I shall be addressing in this lecture.

I mean to be considering a broad family of naturalistic theories of intentionality here; the tension just mentioned does

we ordinarily understand it) requires intentionality (the converse thesis is less obvious). But it is not so clear how closely connected are the problems of rationality and consciousness: can the former exist without the latter? If we find consciousness theoretically daunting (as I argue we should), then we should hope that rationality can be separated from it. There is a general question here: how much of the mind can be explained without being able to explain consciousness? This, as I suggest later, is the same as the question how much of the mind can be explained.

not arise from one sort of theory alone. There are currently a number of theories to choose from: causal theories, functionalist theories, computational theories, teleological theories.<sup>3</sup> Take any of these and ask yourself whether that theory accounts satisfactorily for consciousness: does it, specifically, provide sufficient conditions for being in a conscious state? If it does not, then the question must be faced how it can be an adequate explanation of content *for* conscious states. Consider, for instance, teleological theories (my own favourite). This type of theory identifies the content of a mental state with (roughly) its world-directed biological function. A desire state has a content involving *water*, say, just if that state has the function of getting the organism to obtain water. A perceptual experience represents squareness, say, just if its function is to indicate (covary with) the presence of square things in the environment. But now these contents serve to fix the phenomenological aspects of the states in question, what it is like subjectively to be in them; yet the theory itself seems neutral on the question of consciousness. Certainly the teleological descriptions of the states seem insufficient to confer conscious subjective features on them. *Any* naturalistic theory of the kinds currently available looks to be inadequate as an account of what makes a mental state have a particular *conscious* content, a specific phenomenology. Yet phenomenology seems configured by content.<sup>4</sup>

This question is especially pressing for me, since I have come to hold that it is literally impossible for us to explain how conscious-

<sup>3</sup> For discussions of these approaches see: Jerry Fodor, *Psychosemantics* (The MIT Press, Cambridge, Mass., 1987); Fred Dretske, *Knowledge and the Flow of Information* (The MIT Press, Cambridge, Mass., 1981); Hilary Putnam, *Representation and Reality* (The MIT Press, Cambridge, Mass., 1988); Ruth Millikan, *Language, Thought and Other Biological Categories* (The MIT Press, Cambridge, Mass., 1984); Colin McGinn, *Mental Content* (Basil Blackwell, Oxford, 1989).

<sup>4</sup> My focus in this paper is on the content of perceptual experiences, mental states for which the notion of a subjective phenomenology is best suited. But essentially the same questions arise for thoughts, mental states for which the notion of what it is like to have them seems strained at best (thoughts are not inherently 'qualia-laden'). Thoughts are conscious, of course, and the question, what confers this consciousness, is equally pressing for them as it is for experiences. Moreover, the content of thoughts looks even more closely tied to their conscious features than in the case of experiences; so it is even harder to see how we could pull apart the theory of content for thoughts and the theory of what gives thoughts their conscious aspect. What more is there to the specific way a thought is present in the stream of consciousness than its having the particular content it has?

ness depends upon the brain, even though it does so depend.<sup>5</sup> Yet I also believe (or would like to believe) that it is possible for us to give illuminating accounts of content.<sup>6</sup> Let me briefly explain my reasons for holding that consciousness systematically eludes our understanding. Noam Chomsky distinguishes between what he calls 'problems' and 'mysteries' that confront the student of mind.<sup>7</sup> Call that hopeful student *S*, and suppose *S* to be a normal intelligent human being. Chomsky argues that *S*'s cognitive faculties may be apt for the solution of some kinds of problem but radically inadequate when it comes to others. The world need not in all of its aspects be susceptible of understanding by *S*, though another sort of mind might succeed where *S* constitutionally fails. *S* may exhibit, as I like to say, *cognitive closure* with respect to certain kinds of phenomena: her intellectual powers do not extend to comprehending these phenomena, and this as a matter of principle.<sup>8</sup> When that is so Chomsky says that the phenomena in question will be a perpetual mystery for *S*. He suspects that the nature of free choice is just such a mystery for us, given the way our intellects operate. That problem need not, however, be intrinsically harder or more complex than other problems we can solve; it is just that our cognitive faculties are skewed away from solving it. The structure of a knowing mind determines the scope and limits of its cognitive powers. Being adept at solving one kind of problem does not guarantee explanatory omniscience. Human beings seem remarkably good (surprisingly so) at understanding the workings of the physical world—matter in motion, causal agents in space—but they do far less well when it comes to fathoming their own minds. And why, in evolutionary terms, should they be intellectually equipped to grasp how their minds ultimately operate?

Now I have come to the view that the nature of the dependence of consciousness on the physical world, specifically on the brain, falls into the category of mysteries for us human beings, and possibly for all minds that form their concepts in ways constrained by perception and introspection. Let me just summarize why I think this; a longer treatment would be needed to

<sup>5</sup> See my, 'Can We Solve the Mind-Body Problem?', *Mind*, July 1989.

<sup>6</sup> See my, *Mental Content* (Basil Blackwell, Oxford, 1989). The present paper is an attempt to reconcile the optimism of that book with the pessimism of the paper cited in note 5.

<sup>7</sup> See his, *Reflections on Language* (Pantheon Books, 1975), Chap. 4.

<sup>8</sup> Cf. Fodor's notion of 'epistemic boundedness': *The Modularity of Mind* (The MIT Press, Cambridge, Mass., 1983), Part V.

make the position plausible. Our concepts of the empirical world are fundamentally controlled by the character of our perceptual experience and by the introspective access we enjoy to our own minds. We can, it is true, extend our concepts some distance beyond these starting-points, but we cannot prescind from them entirely (this is the germ of truth Kant recognized in classical empiricism). Thus our concepts of consciousness are constrained by the specific form of our own consciousness, so that we cannot form concepts for quite alien forms of consciousness possessed by other actual and possible creatures.<sup>9</sup> Similarly, our concepts of the body, including the brain, are constrained by the way we perceive these physical objects; we have, in particular, to conceive of them as spatial entities essentially similar to other physical objects in space, however inappropriate this manner of conception may be for understanding how consciousness arises from the brain.<sup>10</sup> But now these two forms of conceptual closure operate to prevent us from arriving at concepts for the property or relation that intelligibly links consciousness to the brain. For, first, we cannot grasp other forms of consciousness, and so we cannot grasp the theory that explains these other forms: that theory must be general, but *we* must always be parochial in our conception of consciousness. It is as if we were trying for a general theory of light but could only grasp the visible part of the spectrum. And, second, it is precisely the perceptually controlled conception of the brain that we have which is so hopeless in making consciousness an intelligible result of brain activity. No property we can ascribe to the brain on the basis of how it strikes us perceptually, however inferential the ascription, seems capable of rendering perspicuous how it is that damp grey tissue can be the crucible from which subjective consciousness emerges fully formed. That is why the feeling is so strong in us that there has to be something *magical* about the mind–brain relation. There must *be* some property of the brain that accounts non-magically for consciousness, since nothing in nature happens by magic, but no

<sup>9</sup> Nagel discusses this in *The View From Nowhere* (Oxford University Press, 1986), Chap. 2.

<sup>10</sup> That is, our natural perception-based sense of similarity underestimates the objective difference there must be between brains and other physical objects, if brains are to be (as they are) the basis of consciousness. To God, brains seem *sui generis*, startlingly different from other physical objects. His sense of similarity, unlike ours, does justice to the uniqueness we know the brain must possess. (Compare the fallibility of our natural sense of similarity with respect to natural kinds.)

form of inference from what we perceive of the brain seems capable of leading us to the property in question. We must therefore be getting a partial view of things. It is as if we were trying to extract psychological properties themselves from our awareness of mere physical objects; or again, trying to get normative concepts from descriptive ones. The problem is not that the brain lacks the right explanatory property; the problem is that this property does not lie along any road we can travel in forming our concepts of the brain. Perception takes us in the wrong direction here. We feel the tug of the occult because our methods of empirical concept formation are geared towards properties of kinds that cannot in principle solve the problem of how consciousness depends upon the brain. The situation is analogous to the following possibility: that the ultimate nature of matter is so different from anything we can encounter by observing the material world that we simply cannot ever come to grasp it. Human sense organs are tuned to certain kinds of properties the world may instantiate, but it may be that the theoretically basic properties are not ones that can be reached by starting from perception and workings outwards; the starting-point may point us in exactly the wrong direction. Human reason is not able to travel unaided in just any theoretical direction, irrespective of its basic input. I think that honest reflection strongly suggests that nothing *we* could ever empirically discover about the brain *could* provide a fully satisfying account of consciousness. We will either find that the properties we encounter are altogether on the wrong track or we shall illicitly project traits of mind into the physical basis.<sup>11</sup> In particular, the essentially spatial conception we have, so suitable for making sense of the nonmental properties of the brain, is inherently incapable of removing the sense of magic we have about the fact

<sup>11</sup> This latter tendency gives rise to illusions of understanding. We think we are seeing how consciousness depends upon the brain when all we are doing is reading consciousness into the physical basis. This tendency is particularly compelling when the brain is conceived as a computer: thinking of neurons as performing computations, we are tempted to credit them with conscious states (or proto-conscious states). Then it seems easy enough to see how neurons could generate consciousness. But, of course, this just pushes the question back (as well as being false): for how do these conscious properties of neurons arise from their physical nature? (Panpsychism now threatens.) If we are to describe physical processes computationally, then we must be clear that this does not involve consciousness—and then it will also be clear that we can't get consciousness out of such descriptions. Either we presuppose what we should be explaining or we find ourselves as far away as ever from our explanandum.

that consciousness depends upon the brain. We need something radically different from this but, given the way we form our concepts, we cannot free ourselves of the conceptions that make the problem look insoluble. Not only, then, is it *possible* that the question of how consciousness arises from the physical world cannot be answered by minds constructed as ours are, but there is also strong positive reason for supposing that this is actually the case. The centuries of failure and bafflement have a deep source: the very nature of our concept-forming capacities. The mind-body problem is a 'mystery' and not merely a 'problem'.

The foregoing is only intended to provide a flavour of the reasons I would give for abject pessimism over the problem of consciousness. My question in this lecture concerns the consequences of such pessimism for the problem of content. Must we suppose likewise that intentionality is closed to our theoretical understanding, that the correct naturalistic theory treats of properties that lie outside the area of reality we can comprehend? Or is there some way to stop the mystery of consciousness spreading to content? Before considering some possible suggestions on how to contain the mystery, let me focus the tension a bit more sharply.

Consider conscious perceptual experiences, such as my now seeing a scarlet sphere against a blue background. We can say, following Thomas Nagel and others, that there is something it is like to have such experiences; they have a subjective aspect.<sup>12</sup> That is to say, there is something it is like *for the subject* of such experiences: subjective aspects of experience involve a reference to the subject undergoing the experience—this is what their *subjectivity* consists in. But we can also say that perceptual experiences have a world-directed aspect: they present the world in a certain way, say as containing a scarlet sphere against a blue background. This is their representational content, what states of affairs they are *as of*. Thus perceptual experiences are Janus-faced: they point outward to the external world but they also present a subjective face to their subject; they are of something other than the subject and they are like something for the subject. But these two faces do not wear different expressions: for what the experience is like is a function of what it is of, and what it is of is a function of what it is like. Told that an experience is as of a scarlet sphere you know what it is like to have it; and if you know

<sup>12</sup> Thomas Nagel, 'What is it Like to be a Bat?', *Mortal Questions* (Cambridge University Press, 1979); Brian Farrell, 'Experience', *Mind*, 1950.

what it is like to have it, then you know how it represents things. The two faces are, as it were, locked together. The subjective and the semantic are chained to each other. But then it seems that any conditions necessary and sufficient for the one aspect will have to be necessary and sufficient for the other. If we discover what gives an experience the (full) content it has, then we will have discovered what gives it its distinctive phenomenology; and the other way about. But now we are threatened with the following contraposition: since we cannot give a theory of consciousness, we cannot give a theory of content, since to give the latter would *be* to give the former (at least in the case of conscious experiences). Accordingly, theories of content are cognitively closed to us: we cannot say in virtue of what an experience has the content it has. Suppose, for example, that we favoured some sort of causal theory of perceptual content: content is fixed by regular causal connexions between experiences and properties instantiated in the surrounding world, say being scarlet or spherical.<sup>13</sup> Such causal facts would be deemed sufficient for having the kind of content in question. But if this content fixes the subjective side of the experience—what it is like for the subject—then we are committed, it seems, to holding that such causal facts are sufficient for this subjective side also. For what fixes content fixes qualia. But these causal conditions seem manifestly insufficient for subjectivity, intuitively, and the claim contradicts the closure I said I concede. Intentionality has a first-person aspect, and this seems impossible to capture in the naturalistic terms favoured by causal theories and their ilk.<sup>14</sup> If consciousness is a mystery, then so must its content be. So the challenge runs.

How, if at all, can we escape this argument? One response would be not to try: accept that intentionality is inexplicable by us but insist that it is not inherently mysterious or inconsistent with what we know of the physical world. This would be to extend to content the treatment I would propose for consciousness. About consciousness I would say that there is no objective

<sup>13</sup> This kind of theory is defended by (among others) Tyler Burge, 'Individualism and Psychology', *Philosophical Review*, January 1986. I criticize such views in *Mental Content* (Basil Blackwell, Oxford, 1989).

<sup>14</sup> Such theories stress the third-person perspective: how we determine what someone else is referring to or thinking about. But we must not forget the perspective of the subject: how he experiences the intentional directedness of his mental states. It is the same stress on the third-person perspective that makes the likes of functionalism about sensations seem more adequate than it ought to seem.



miracle in how it arises from the brain; it only seems to us that there is because of the veil imposed by cognitive closure. We project our own limitations on to nature, thus making nature appear to contain supernatural facts. In reality, there is no metaphysical mind-body problem; there is no *ontological* anomaly, only an epistemic hiatus. The psychophysical nexus is no more intrinsically mysterious than any other causal nexus in the body, though it will always strike *us* as mysterious. This is what we can call a 'nonconstructive' solution to the problem of how consciousness is possible. But if that solution removes the basic philosophical problem, as I think it does, then we can say the same about intentionality. We do not need to be able to produce a constructive solution to 'Brentano's problem' in order to convince ourselves that there is no inherent mystery in the phenomenon of intentionality; we can rest secure in the knowledge that *some* property of the physical world explains that phenomenon in an entirely natural way—though we cannot ever discover what that property is.<sup>15</sup> To the omniscient intellect of God intentionality in a physical organism is no more remarkable than digestion is. Thus there is no pressure towards eliminativism about content arising from the fact that *we* can never make content (physically) comprehensible to ourselves; any more than a minded creature who is constitutionally unable to grasp the correct theory of digestion has to deny that anything ever gets digested. So we can, according to this response, solve the philosophical problem of intentionality without actually *specifying* the correct theory.

I do not think this nonconstructive response can be rejected on general grounds, since I believe it applies to the case of consciousness. But I think it is implausibly extreme in the case of content; for we can, I believe, produce naturalistic theories of content

<sup>15</sup> Here, then, is a possible response to Hartry Field's demand that truth and reference be reducible if they are to be respectable: see his 'Mental Representation', in Ned Block (ed.), *Readings in Philosophy of Psychology* (Harvard University Press, Cambridge, Mass., 1981). We need to distinguish being able to *give* a reduction from knowing that a reduction *exists*—in order not to rule out the possibility that the reduction can be specified only in a science that is cognitively inaccessible to us. We cannot infer elimination from irreducibility *by us*. Nor can we simply *assume* that the correct naturalistic account of intentionality employs 'broadly physical' notions, if this means that these notions do not extend our present physical concepts beyond what is intelligible to us. In a word, we must not be dogmatic conceptual conservatives. The correct reduction (if that is the right word) might not be recognizable by us as correct. (I take this to be an expression of realism.)

that provide substantial illumination as to its workings. It is not as if the theories now available strike us as just hopelessly misguided, as telling us nothing whatever about the nature of intentionality. Whereas I do think that the usual theories of consciousness (e.g. functionalism) do not even begin to make a dent in our incomprehension. Thus it seems to me that teleological theories, in particular, promise to shed a good deal of light on the roots of intentionality; they provide real insight. Who can deny that the vast amount of work devoted to the nature of reference and belief in the last twenty or so years has added significantly to our understanding of their nature? Something, I venture to suggest, has been learned. So it cannot be that the whole nature of intentionality is hidden from us, that we simply cannot form the kinds of concepts that would shed light on it. The question is how to square this apparent illumination with extreme pessimism about consciousness. How is such illumination *possible*, given that we are completely in the dark about consciousness?

At this point it is natural to pin one's hopes on what I shall call the 'insulation strategy'. The insulation strategy proposes radically to separate the two theories: in particular, it proposes to do the theory of content in complete isolation from the theory of consciousness. How might this insulation of theories be made plausible? The obvious first move is to switch theoretical attention to (so-called) subpersonal content, the kind that exists without benefit of consciousness. We attribute content of a sort to machines and to subconscious processes in the nervous system; and this kind of content might be thought to be explicable without bringing in consciousness. It is true that content is also possessed by conscious states, but this is only a contingent truth about content, a dispensable accretion. Then once we have a theory for subpersonal content we can extend it to conscious content, simply by adding in the fact that the content is conscious. In principle, this strategy insists, the conditions necessary and sufficient for content are *neutral* on the question whether the bearer of the content happens to be a conscious state. Indeed, the very same range of contents that are possessed by conscious creatures could be possessed by creatures without a trace of consciousness. Consciousness is simply a further fact, super-added; it is not itself in any way constitutive of content. This contingency claim might then be bolstered by the consideration that the outstanding problem in the naturalistic theory of content—namely, accounting for the possibility of error or

misrepresentation—does not seem to require invoking consciousness: it is not the fact that a state is conscious that makes it susceptible to error and hence semantic evaluation. We do not ascend from mere natural indication or nomic dependence to full-blown truth and falsity by ensuring that there is something it is like to be in the state in question. Subjectivity is not what creates the possibility of error. Hence subjective features lie quite outside the proper domain of the theory of content.

There are two problems with this suggestion. The first is tactical: we do not want the possibility of a theory of content to depend upon the particular conception of the relation between content and consciousness that the suggestion assumes. One view, by no means absurd, is that *all* content is originally of conscious states. There is no (underivative) intentionality without consciousness. (Brentano's thesis was that all consciousness is intentional; this 'converse Brentano thesis' is that all intentionality is conscious—or somehow derivative from consciousness.) Our attributions of content to machines and cerebral processes is, on this view, dependent or metaphorical or instrumental; there would be no content in a world without consciousness. Accordingly, we labour under an illusion if we think we can *complete* the theory of content without even mentioning that contentful states are associated with consciousness. There is no ofness without likeness. When we think we are conceiving of content in the absence of consciousness we are really treating a system *as if* it were conscious, while simultaneously denying that this is what we are up to.

Now it is not that I myself agree with this extreme thesis of dependence; I have yet to see a convincing argument for the claim that any kind of representation worthy of the name requires consciousness. But I would agree that the possibility of subpersonal content of *some* kind does not serve to insulate the two theories when it comes to the kind of content distinctively possessed by conscious states. And this brings us to the second point. There may indeed be two species of content, personal and subpersonal, but this does not show that the personal kind lacks distinctive properties that tie it essentially to consciousness. I doubt that the self-same *kind* of content possessed by a conscious perceptual experience, say, could be possessed independently of consciousness; such content seems essentially conscious, shot through with subjectivity. This is because of the Janus-faced character of conscious content: it involves presence to the subject, and hence a subjective point of view. Remove the inward-looking

face and you remove something integral—what the world *seems* like to the subject. Just as there are two types of ‘meaning’, natural and non-natural, so there seem to be two types of content, conscious and nonconscious; the subjective perspective creates, as it were, a new and special kind of content. This is why what an experience is as of already contains a phenomenological fact—how the subject is struck in having the experience. So we cannot hope to devise an exhaustive theory of the nature of conscious content while remaining neutral on whether such content is conscious. Content distinctions confer subjective distinctions. Experiential content is essentially phenomenological.

I suspect that the insulation strategy is fuelled by a conception of consciousness that we can call the ‘medium conception’: consciousness is to its content what a medium of representation is to the message it conveys. Compare sentences, spoken or written. On the one hand, there is their sound or shape (the medium); on the other, their meaning, the proposition they express. We can readily envisage separate studies of these two properties of a sentence, neither presupposing the other. In particular, we could have a theory of the content of sentences that was neutral as to their sound or shape. The meaning could vary while the sound or shape stayed constant, and there could be variations in sound or shape unaccompanied by variations in meaning. Message and medium can vary along independent dimensions. Suppose, then, that we try to think of perceptual experience in this way: subjective features are analogous to the sound or shape of the sentence, content to its meaning. The content is *expressed* in a particular conscious medium but we can in principle separate the properties of the medium from the message it carries. What it is like to have the experience is thus fixed by intrinsic features of the medium, whereas what the experience is about is fixed by certain extrinsic relations to the world. According to this conception, then, the absolute intractability of consciousness need not infect the theory of content in the slightest. Consciousness is to be conceived, in effect, as a mysterious medium in which something relatively mundane is (contingently) embedded.

I think the medium conception is the kind of view which, once clearly articulated, sheds whatever attractions it may have initially possessed. In effect, it tries to treat perceptual experience as if its phenomenology were analogous to that of (non-representational) bodily sensations: content comes from subtending this intrinsic phenomenology with causal or other relations to the

world, these relations being strictly orthogonal to that intrinsic phenomenology. Or again, it tries to conceive of experiential content as if it operated like truth or veridicality: whether a belief is true or an experience veridical is not a phenomenological property of the state in question, so that any theory of what confers these properties need not encroach on consciousness itself. A causal account of veridicality, for example, is not, and is not intended as, an account of what gives an experience the representational content it has (what it is as of). *If* we could think of content itself as lying in this way 'outside' of phenomenology, then we could indeed insulate the two theories. But, as I have insisted, this attempted extrusion of the subjective from the semantic just does not work. The content of an experience simply does contribute to what it is like to have it, and indeed it is not at all clear that anything else does. A visual experience, for example, presents the world to the subject in specific ways, as containing spatially disposed objects of various shapes and colours, and this kind of 'presentation-to' is constitutive of what it is like to have visual experience. It is true, of course, that different sense-modalities may present the same kinds of environmental feature, e.g. shape or texture—as with sight and touch—but the subjectively distinct experiences that present these features also present *other* features. It is not that sight and touch present precisely the *same* range of features yet differ phenomenologically, so that we need something like a medium conception to capture the difference; it is rather that they overlap in the features they present at certain points but are disjoint at others—notably, in the secondary qualities they present. These differences in the range of contents available to different types of experience seem enough to capture the obvious phenomenological differences in the experiences associated with different senses. Bats perceive different secondary qualities from us when they employ their echolocation sense; it is not that they perceive precisely the same qualities and embed them in a different (non-representational) medium. But even if there were subjective distinctions that could not be captured in terms of distinctions of content, this would not help the insulation strategy, since there are too many subjective distinctions that *are* generated by distinctions of content. The difference between a visual experience of red and a visual experience of green just is a difference in what it is like to have these two types of experience. The case is quite unlike the difference between a veridical and an hallucinatory

experience, or a true belief and a false one. Content, we might say, is *internal* to phenomenology; the link here is anything but contingent.

If this is right, then we cannot suppose that the theory of content simply has nothing to do with the nature or constitution of consciousness. Since distinctions of content can constitute (or contribute towards) distinctions of phenomenology, we cannot totally insulate the theory of the former from the theory of the latter; we must admit that a correct theory of content will deliver resources sufficient to capture subjective features of conscious states. But if we are convinced that no naturalistic theory of the kinds available to us can explain conscious features, then we are again in a state of tension. Either we can explain features of consciousness ('qualia') naturalistically or we can't explain content naturalistically. The fate of the one theory seems yoked to the fate of the other. Yet I, for one, would like to believe that we can make progress with content, while accepting that consciousness is beyond us. Where then can I turn to have this tension relieved?

Instead of attempting to insulate the two theories entirely, I want to suggest that we limit the scope of the theory of content. We should accept that there is a part or aspect of intentionality that our theories do not and probably cannot capture, but we should also hold that there is a part or aspect that they do have some prospect of illuminating. There is *partial* cognitive closure with respect to content: we can naturalize certain properties of the total phenomenon but we cannot naturalize all of its properties (though, as I said earlier, all properties are in themselves entirely natural). And this will imply that there are *some* features of consciousness—subjective features—that we can treat naturalistically. There is a feasible branch of the theory of content that delivers an account of certain phenomenological facts: but this falls short of a full explanation of conscious intentionality.

Let me distinguish two questions. The first is the question what *individuates* contents: what accounts for identity and difference between contents, what makes a content of this rather than that. We classify experiences according to what they represent, and the question is what principles underlie these classifications. The second question concerns the *nature* of content: what it consists in for a creature to have intentional states at all, what makes a creature enjoy mental 'directedness' on to the world in the first place. Thus, we can ask what natural facts make a creature an intentional being, and then we can ask what natural facts *target*

this intentionality in specific ways. The question of nature is the more fundamental question: it asks what this directedness, grasping, apprehension, encompassing, reaching out ultimately consists in. It wants to know by virtue of what natural facts the mind is endowed with the power to 'point' beyond itself. The question of individuation takes this for granted and enquires how the intentional capacity picks up the particular objects and properties it does. *Given* that consciousness has the power to 'lasso' things in the world, what determines the direction of its throw? Putting it in terms of linguistic intentionality or reference: we can ask what makes a physical organism capable of referring (the act itself), and we can ask how it is that this act is tied down to particular objects and properties. 'What is reference?' is one question; 'How does reference get targeted this way rather than that?' is another question.

Now, assuming this distinction is sufficiently clear, I can state my proposal: the *nature* of intentionality is cognitively closed to us but the *individuation* of intentional contents is in principle open. We can say what makes a content of this rather than that but we cannot say what the relation of intentionality itself consists in. We cannot specify, in naturalistic (i.e. broadly physical) terms, the essential nature of the conscious mental act of apprehending states of affairs, but we can say in such terms what distinguishes one such act from another. Let me now try to defend this proposal. First I will explain why the proposal is consistent. Then I will defend the pessimistic part of the proposal. Finally I will urge a qualified optimism about the question of content individuation.

The proposal is consistent because we do not need to fathom the nature of the intentional *act* in order to provide constraints on the identity conditions of instances of the act. I can tell you what distinguishes referring to redness from referring to greenness without being able to tell you what referring is *au fond*. The direction of reference may be constrained by relations with which reference itself cannot literally be *identified*. An analogy from action theory may help here. We can ask what distinguishes different kinds of world-directed bodily action without asking what the nature of intentional action in general is. Thus I can tell you what distinguishes intentionally kicking a brick from intentionally kicking a cat—there are different objects on the end of my toe—without having to explain what intentional action is in general. Consider, then, causal theories of mental aboutness. I can tell you, in terms of causal history, what distinguishes

thinking about London from thinking about New York—there are different cities at the causal origin of these thoughts—without having to venture on the question what mental aboutness is to start with. The causal relations in question make these thoughts home in on certain objects, but we do not need to infer that mental aboutness is reducible to these relations. I don't have to be able to explain or analyse the act of grasping itself in order to be able to lay down laws that fix what is grasped. I don't have to be able to provide a naturalistic account of the intentional *structure* of consciousness in order to be in a position to pin down what gives that structure the specific content it has. Specific content is, as it were, the 'logical product' of the intentional capacity and the natural relations that target that capacity in particular ways; the capacity is not reducible to the relations. In view of this distinction of questions, we have to be very careful when we offer what we are pleased to call a 'theory of intentionality/reference'. Suppose we favour causal theories of perceptual content: content is individuated by regular causal links between experiences and properties instantiated in the subject's environment. It is tempting to suggest that such links give us the very nature of perceptual representation, that the conscious act of enjoying an experience as of a scarlet sphere against a blue background is analysable as a special kind of causal relation. But if I am right this is not what we should say. Rather, we should say that causal relations tie the intentional structure of perceptual experience down to specific states of affairs but that such relations do not constitute the very nature of that structure. Intentional directedness is not exhaustively analysable as a causal relation, however complex. And similarly for teleological theories. Neither do we need to suppose this in order to find a point for naturalistic theories of content; we need rather to locate their legitimate area of application some way short of a full account of what it is to stand in intentional relations to things.

The pessimism about the essential nature of intentionality can be motivated in two ways. First we can simply deduce it from pessimism about consciousness: if consciousness cannot be explained (by us) naturalistically, in broadly physical terms, then neither can the constitutive structures of consciousness. The intentionality of experiences and thoughts belongs with the subjective 'feel' of sensations: neither admits of objective physical explanation. But, second, we can also generate a mood of pessimism more directly: we can ask ourselves whether it really seems plausible that any of the standard theories capture the



complete nature of conscious intentionality. In the case of sensations we have a strong sense that standard naturalistic theories, e.g. reductive functionalism, omit something essential—the ‘feel’ of the sensation. And I think our intuitions about intentionality parallel our intuitions about sensations: it really does seem that causal or teleological theories omit something essential in the intentional relation as it occurs in consciousness. They do not capture that phenomenological feature we describe (somewhat metaphorically) as grasping, apprehending, reaching out, taking in, and so forth. There is an *internality* about the relation between an experience and its object that seems hard to replicate in terms of ‘external’ causal or teleological relations. Presence *to* the subject of the object of his experience seems not exhaustively explicable in terms of such natural relations. These kinds of relations hold, after all, between all sorts of things, not just brains and items in their environment, and it seems unsatisfactory to try to assimilate conscious intentional directedness to these ordinary relations. Conscious intentionality is more special than this sort of account suggests. (This is, of course, why Brentano claimed that intentionality is what distinguishes minds from mere physical objects.) Naturalistic theories fail to do justice to the *uniqueness* of conscious intentionality. Nothing we know about the brain, including its relations to the world, seems capable of rendering unmysterious the capacity of conscious states to ‘encompass’ external states of affairs.<sup>16</sup> I think this is a

<sup>16</sup> Two thinkers who have recognized the mysterious-seeming nature of meaning and reference are Thomas Nagel and Ludwig Wittgenstein. Nagel draws attention to the way meaning seems to be able to ‘take in’ much more of the world than its basis in the particular doings and undergoings of speakers could permit: it can reach across vast stretches of space and time; it has a universality or generality that transcends the particular actions and experiences of speakers; it determines indefinitely many uses of language, past and future, as correct or incorrect. See his, *What Does It All Mean?* (Oxford University Press, 1987), Chap. 5. Wittgenstein, for his part, speaks of ‘the mysterious relation of the object and its name’, and he says of the ‘mental activities’ of wishing and believing that ‘for the same reason [they] have something mysterious and inexplicable about them’: *The Blue and Brown Books* (Basil Blackwell, Oxford, 1958), pp. 172–3. Wittgenstein’s idea, though, is that this sense of mystery arises from a (correctable) mistake: ‘A primitive philosophy condenses the whole usage of the name into the idea of a relation, which thereby becomes a mysterious relation’ (p. 173). I am inclined to agree with him about the aura of mystery, but I doubt that it can be dispelled in the way he suggests, namely by reminding ourselves of how we actually use names or ascribe propositional attitudes. I don’t think a deflationary response of this kind is adequate to the problem.

very primitive intuition, by which I suspect many of us have been struck at some point in our philosophical lives. How *can* our minds reach out to the objects of experience? What is it about our brains, and their location in the world, that could possibly explain the way consciousness *arcs out* into the world? Consciousness seems to extend an invisible hand into the world it represents (if I may put it so): how on earth could my *brain* make that possible? No ethereal prehensile organ protrudes from my skull! Phenomenologically, we feel that the mind 'lays hold' of things out there, mentally 'grasps' them, but we have no physical model of what this might consist in. We flounder in similes. It is precisely our perplexity about this question that makes it seem to us that there could be a creature whose brain had all the same natural properties and relations as ours and yet enjoyed no such conscious arcing out. For none of the natural properties and relations we come across seems to add up to what we know from the first-person point of view of conscious aboutness. It is thus reasonable to suspect that cognitive closure is operative here. Somehow we are not keyed in to the kinds of natural fact that actually underlie intentionality—as we are not to consciousness in general. Something about our make-up explains how consciousness can reach out into the world in the way it does, but we seem constitutionally blind to what that something is.

Cautious optimism is possible, however, since we do not need to explain everything about intentionality in order to be able to say something illuminating about it. And I think it is undeniable that illuminating things have been said about content in recent years; all is not darkness. Teleological theories, in particular, seem to me to contain valuable insights. The question is *what* precisely has been illuminated. And my suggestion is that these naturalistic theories should be seen as contributions to the individuation conditions of mental states: they tell us what differentiates one kind of intentional state from another: they tell us how intentional states collect their specific content. They may also tell us something about the natural antecedents of conscious intentionality—what basic natural relations got transformed by consciousness into genuine content. First there were preconscious states with certain functions relating them to things in the world; then consciousness built upon this natural foundation to produce the intentional relation. The 'intentional arc' is not reducible to this foundation but it takes its rise from it. So there is room for naturalistic speculation about where intentionality came from, if not what it ultimately consists in. We can pursue these more

modest questions without having to take on the full explanatory task of reducing intentionality to something we can understand, something broadly physical. In fact, something like this perspective is already implicit in much work on reference and content. It is not invariably assumed that causal theories (say) give us the real nature of the reference relation, that they successfully analyse the capacity to refer; rather, they tell us how that capacity gets targeted, what constrains the direction of acts of reference.<sup>17</sup> So we can be grateful for this kind of illumination without insisting that it be spread across the whole phenomenon.

Yet there is a residual puzzle. We have resisted the insulation strategy, arguing that content colours consciousness. Differences of content do determine differences of subjectivity; 'ofness' fixes 'likeness'. But this staining of subjectivity by reference does imply that we can provide a naturalistic theory of subjective distinctions, since we can say in naturalistic terms what individuates the content of experience. Here we have an objective handle on to the constitution of the subjective. An experience as of a red square thing is subjectively distinct from an experience as of a green triangular thing, in virtue of the fact that different kinds of objects are represented; and this distinction can be captured, we have agreed, in terms of natural relations that these experiences stand in to the properties represented—say, teleological relations. So it looks as though we are committed to accounting for *some* features of consciousness naturalistically; not *all* phenomenological facts are closed to us. I think this does indeed follow: there are some features of consciousness whose natural explanation, in broadly physical terms, is in principle available to us. Our

<sup>17</sup> This seems the right way to interpret Saul Kripke's remarks about naming and reference in *Naming and Necessity* (Basil Blackwell, Oxford, 1980). Kripke disavows any intention of analysing or reducing the relation of reference, offering us only a 'picture' of how reference operates; but he does give us substantive constraints on *which* object is being referred to by the use of a name—the object which lies at the origin of the 'causal chain' of uses that historically lead up to the use in question. And there is nothing in the kind of closure I acknowledge to preclude descriptive work in the theory of reference: distinguishing the different kinds of referential device, articulating the modes of identification that underlie uses of these different devices, showing how sense and reference are related in different cases, and so forth. Nothing I say undermines the viability and usefulness of, say, Gareth Evans's work in *The Varieties of Reference* (Oxford University Press, 1982). What I am doubting is the possibility of a certain kind of explanatory enterprise: giving a broadly physical account of the very nature of the reference relation. We can prune the pretensions of causal theories (say) without declaring them completely out of a job.

concept-forming capacities afford us partial access to the natural basis of these subjective features of consciousness. But this is puzzling because one would expect the closure to be total: how can it be impossible for us to explain how consciousness arises from the physical world and yet not so very difficult to account naturalistically for distinctions *within* consciousness? Why should the general phenomenon of consciousness be so recalcitrant to natural explanation while specific determinations of consciousness yield to naturalistic account? It's puzzling. Even where consciousness is not mysterious it is mysterious why it is not mysterious!

This puzzle should be set beside another. A moderate externalist about content will hold that objective properties, e.g. being square, enter into the identity of contentful states; they occur as 'constituents' of content.<sup>18</sup> Thus objective properties penetrate experiences in ways that fix their phenomenology. Again, the subjective is invaded by the objective. Combining this act of colonization with the previous one we get a double dependence of the subjective on the objective: objective items figure as 'constituents' of subjective states, so shaping their phenomenology, *and* these states collect those objective 'constituents' by way of objective natural relations—say, biological function. What now begins to look mysterious is the way consciousness is so resistant to objective physical reduction and yet is so permeated by the objective and physical. Consciousness, as it were, appropriates the objective while holding itself aloft from it; it takes the physical in but it refuses to be ruled by it. And, oddly enough, it is just this capacity to 'incorporate' the physically objective, to bring it within consciousness, that the physical brain seems so inadequate to.<sup>19</sup> The puzzles multiply. But then the more you

<sup>18</sup> For a discussion of this see my *Mental Content* (Basil Blackwell, Oxford, 1989). I set aside here the question of how secondary qualities enter into the content of experience. If these are subjectively constituted, then there is a sense in which the subjective gets turned back on itself when colours (say) penetrate the content of colour experience. Still, colours are properties of external objects, so colour experience—like shape experience—does reach out to the world beyond the subject. (We may wonder whether the ultimate explanation of why we perceive secondary qualities at all is one of those questions about consciousness whose answer is forever closed to us. That would certainly account for my struggles to explain it in *The Subjective View*: Oxford University Press, 1983.)

<sup>19</sup> Genuine externalism therefore requires us to reject the more obvious kinds of physicalism, since the brain cannot incorporate the external in the way the mind can. We have no physical model of how consciousness can lay hold of the physical world in the peculiar way it does.

think about consciousness the more puzzling it comes to seem. It is comforting to reflect that from God's point of view, i.e. the point of view of Nature, there is no inherent mystery about consciousness at all. The impression of mystery derives from our own incurable cognitive poverty, not from the objective world in which consciousness exists. There is no real magic in the link between mind and matter, however incapable we are of seeing how the trick is done. Cold comfort, perhaps, but whoever said that the nature of the mind should be fully accessible to those with a mind?<sup>20</sup>

<sup>20</sup> I am grateful for comments to Thomas Nagel, Simon Blackburn, and various members of an Oxford discussion group.