

Real Patterns and Surface Metaphysics

§1. The Naturalist Imperative.

Naturalism is supposed to be a Good Thing. So good in fact that everybody wants to be a naturalist, no matter what their views might be¹. Thus there is some confusion about what, exactly, naturalism is. In what follows, I am going to be pretty much, though not exclusively, concerned with the topics of intentionality and consciousness, which only deepens the confusion for these are two areas – perhaps the *last* areas – where it remains possible to doubt the virtues of a naturalistic treatment.

If taken as an expression of the urge to avoid belief in the non-existent and the false, who would deny the virtue of naturalism? But if the non-existent and the false are frankly magical entities like the vital spirit, Cartesian consciousnesses, immaterial bearers of meaning or a supernatural source of intrinsic intentionality, how strong a bulwark is required to save us from error? Does not the consistent trend across more than three hundred years of modern scientific investigation provide us with sufficient evidence to ensure, at least, the supervenience of the phenomena of life, meaning, intentionality and consciousness upon the natural processes of the world? Although the precise grounds and details of these supervenience relations remain very uncertain, it cannot be seriously doubted that all these (along with innumerable still less controversial examples) phenomena are at bottom ‘fully natural’.

Naturalism expresses more than a faith in, but also the desire to *enter into* the orderly community of the *real* sciences. This religious feeling comes in familiar varieties: at one extreme, the fundamentalist Unitarian is remembered for the doctrine of the Unity of Science, which espoused the outright reduction of field of knowledge to physics, reserving for all that resisted reduction the ontological hell of non-existence. At the other extreme we find the new-age liberal theology of *mere* supervenience, unaccompanied by any attempt at reductive analysis, whose Hell is the hell of vacuity and quietism.

Between these extremes falls a more hopeful and optimistically temperate naturalism, which requires neither outright reduction to fundamental science nor yet an empty faith in the naturalist outlook. The rules of this naturalism are straightforward and remain at heart reductionist in spirit: they require us to provide an explanation of the target phenomenon (be it intentionality, consciousness or plain old chemistry) in terms of *Something Else* which is irreproachably ‘natural’ and which does not itself appeal to or depend upon features of the target. The rules of this hopeful naturalism can be codified as follows.

The Rules

X has been *naturalised* iff

- (1) X has been explained in terms of Something Else.
- (2) The Something Else does not essentially involve X.
- (3) The Something Else is *properly* natural.

This notion of naturalization has several virtues. It is reasonably clear and is directly and

¹ For example, John McDowell denies that he is a physicalist but nonetheless regards his view of mind as naturalistic. David Chalmers goes so far as to label his own view ‘naturalistic dualism’!

quite properly aimed at the *scientific* integration of the naturalizer's targets. After all, it would seem very strange if not perverse first to embrace the scientific view of the world and then boast about how there are some phenomena that defy all attempts to give an account of how they fit into that world-view. In terms of the idea of supervenience, the need for an explication of the supervenience relation between target and base domains is obvious (otherwise, to generalize a remark of Simon Blackburn's supervenience is part of the problem rather than part of the solution). Still, there is no guarantee that such explications must be forthcoming and I'll label any view which *denies* the possibility of a Rule-based naturalization (for some domain) *mysterianism*².

The Rules evidently comport well with the several successful naturalizations we already have in hand. Consider chemistry. Although there is no prospect of a full-fledged reduction of chemistry to physics, it is pretty clear that quantum mechanics has succeeded in naturalizing chemistry according to the Rules. Chemical interactions are explicable in physical terms, not in their full detail, but in their very *nature*. I believe that the process of heredity has been (or is very close to being) similarly naturalized by the bio-chemical understanding of genetics. And while no one knows how life originated on Earth (let alone anywhere else) the phenomenon of life certainly seems to be a reasonable goal for a naturalization which will likely be attained quite soon.

It is not only scientific domains that have been naturalized; 'weather', 'storm', 'rain' are hardly scientific terms, but the weather has been explicated by the study of fluid dynamics and it is only an – unfortunately inevitable and over-hyped – inability to collect and process enough data that prevents accurate and reasonably long term weather prediction. This list could be long extended, but as we ascend towards more complex phenomena we find ourselves forced to say that we see 'how naturalization would go' rather than being able to trot out a triumphant naturalization. Allowing for this weaker sense of naturalization, it seems to me that now, at the end of the 20th century, it is *only* in the case of the mind (or certain of its features) that we really lack a good sense of how naturalization should go³.

² The name is borrowed from Owen Flanagan's characterization of a well known particular instance of such a view.

³ Perhaps a caveat or two are in order here. It is among the things that *ought* to be naturalized that the mind stands as, at present, especially problematic. It is not clear to me that we need to worry about naturalizing numbers, and other abstractions, though it does seem that all these things can probably be understood in terms of minds and the thoughts they contain. Roughly speaking, what ought to be naturalized are the things that 'push and shove'; those things that make things happen in the world. If numbers, for example, should somehow fall into this category we'll just have to embrace supernaturalism and find the world a *much* stranger place than we thought. (But even here, wouldn't the oddity of the imagined situation come down to how *minds* come into contact with numbers and the like?) On the other hand, metaphysics won't let go here. Isn't it *properties* that cause things, and aren't properties abstract objects? I would say rather that it is things having properties (that is, more or less, *events*) that cause, and property instances – at least those that make things happen – are just the sort of thing that ought to be naturalized. Outside the realm of

Nonetheless, the attempt to present naturalist theories of mind goes back a long way, at least to Plato. Recall in the *Theaetetus* Plato's analogical gropings towards a theory of knowledge, which can easily be seen as an early effort to naturalise intentionality, in the relevant sense of following The Rules. You will remember that Plato likens mental content, under the guise of knowledge, first to impressions in a block of wax and then to a set of captive birds which remain free to fly about the aviary of the mind. Such models unfortunately end up breaking The Rules. For example, the birds can stand in for mental contents only if there is some way to recognise them and this notion of recognition clearly threatens to be viciously intentional. That is, as Plato explains, we can't account for the difference between one of these natural items expressing or signifying a truth as opposed to a falsehood (what Plato calls 'knowledge' and 'ignorance' respectively) unless some intermediate step of recognising the import of these items falls between the occurrence of the item and its significance. But of such an intermediate step Plato wisely notes, in the voice of a 'destructive critic': 'are you going to tell me that there are yet further pieces of knowledge about your pieces of knowledge and ignorance, and that their owner keeps these shut up in yet another of your ridiculous aviaries or waxen blocks ... On that showing you will find yourselves perpetually driven round in a circle and never getting any further' (Plato, *Theaetetus*, 200 b-c)). In honour of this early failure, let's call the violation of Rule 2, Plato's problem.

Many attempts at naturalization have been accused of succumbing to Plato's problem. To see how this goes in a modern setting, I want briefly to consider two non-trivial examples.

The first is Ruth Millikan's theory of intentionality, sometimes called bio-semantics. The theory depends upon an account of *functions* which, very roughly, sees the function of X as what provides the *explanation* of why X's have been 'reproductively successful' within a population (very abstractly conceived of course; this account goes far beyond familiar biological functions). Intentionality is then understood in terms of the functions of 'symbols' to carry information. Of course, lots of things – potential symbols, so to speak – are reliably hooked up to features of the world (smoke and fire, to take a common example) but only for those hookups in which it is the *function* of the candidate symbol to carry information will there be any intentionality or meaning. A little more precisely:

Y means X if X is the basic factor explaining (in an evolutionary or quasi-evolutionary way) the continued existence and historical proliferation of *interpretations* of Y's (by the relevant interpreting 'devices').

Such explanations are what Millikan calls Normal Explanations and pick out what she labels the Proper Function of the sign, or, in her own inimitable prose: 'the most dominant notion of what is signed by signs is derived by reference to the *direct* proper function of these signs themselves, hence to resulting adapted proper functions of interpreting devices qua taking these signs as immediate adaptors' (1984, p. 43).

An example will make this clearer (I borrow the example from Robert Cummins's discussion of Millikan's theory in Cummins 1989): In the famous 'dance' of the honey bees the

the abstract, the ultimate source of existence itself must remain mysterious, short of a physics that emerges out of pure mathematics. The naturalist rests content with finding the ground of all things in the presumed fundamental structure of the world.

orientation of the dance is a sign of where nectar-bearing flowers are. Why? Because the interpretation of the dance by other bees as indicating the presence of such flowers in the appropriate direction explains why the dance proliferated and continues into the present. It may be objected that a huge range of factors actually helped to establish the bee dance. True, but it was the presence of flowers that the dance was ‘selected for’⁴ not, say, the absence of predators in a sufficiently large number of cases where bees were inclined to forage after observing a sister’s dance, or, equivalently, it is not by appeal to the past absence of predators that one can make clear why the dance persisted and proliferated.

Note, by the way, that the reference to interpreting devices prevents just anything with a proper function from becoming a sign – that is, it blocks a version of pansemanticism which might threaten Millikan’s view with vacuity. For example, the heart has as its proper function the pumping of blood and the word ‘red’ has as its proper function the indicating of red but only the latter operates through and essentially through interpreters. That is, whereas the explanation of the proliferation of sign devices needs to make reference to the interaction between sign producers and ‘consumers’ there is no such need in explanations of the proliferation of features that possess ‘non-semantic’ proper functions. Of course, the idea that *interpretation* is an essential feature of the operation of signs awkwardly recalls the difficulty Plato found in the wax-block and caged-birds models of thought. In general the notion of interpreter device must itself be given an entirely non-intentional explanation, most likely in terms of the behaviour that helped, and continues to help, to ‘fix’ the sign into the interpreters’ world. This is not a trivial task but according to The Rules, the complete naturalization of intentionality requires a non-intentional treatment of interpreters. I am sure in practice this condition will be extremely difficult to fulfil, for it amounts to no less than providing a non-cognitive theory of the interpreters’ sign-response behaviour. It strikes me that if we could provide such accounts we would not be just one step closer to naturalizing intentionality, but would have already succeeded in naturalizing intentionality. In the case of the bees, it is of course tempting to suppose that they respond to their sisters’ dances in a ‘mechanical’ way without the intervention of beliefs (or other cognitive states) about or genuine interpretations of these dances, but no one can yet claim to understand the basis of such a mechanism.

It has been argued however that bio-semantics – in common with all the other theories on offer – may face difficulties solving a core problem facing any account of intentionality, namely the problem of assigning the *properly specific* information to particular symbols. Crudely speaking, the problem is to distinguish the proper meaning of ‘horse’, i.e. *horses* from other possible meanings that seem identically capable of reproducing themselves within a reproductive family of symbol systems, for example, *horse or anything indistinguishable from a horse to 20th century science*. As the growing intricacy of the attempts to solve it attest, this problem is deeper than it looks. But as we’ve seen bio-semantics has a ready response to it which is quite plausible. It may be that the reproducing symbol system couldn’t care less about the difference between these two possible meanings since there is nothing in the use of the symbol ‘horse’ that could

⁴ For the distinction between ‘selection for’ and ‘selection of’ see Sober (1985), pp. 97ff. Please note that Plato’s problem looms already – *selection for* is an intensional notion and it is very unclear that appeal to it could be a legitimate part of a naturalization of intentionality.

have distinguished them, but it is ludicrous to claim that the two meanings are equally good *explanations* of why ‘horse’ is connected to what it is connected to (save for any who happen to believe that there are, and have been, clever alien or robot horses interspersed with the genuine article – but that is a *ludicrous* belief).

Now Plato’s problem reappears. On this view of bio-semantics, it is impossible to understand how symbols connect to the world without understanding what *explanation* is; the notion of explanation is a key component of the workings of the theory. It is unfortunately all too clear that the notion of explanation is implicitly intentional. We could caricature the theory something as follows: Symbol S means O just in case the proper explanation of S’s role involves appeal to S’s carrying the information that O. Since S carries lots of information besides O we need some mechanism to narrow down the field. But if the mechanism explicitly appeals to what is a ‘good explanation’ then we have succumbed to Plato’s problem. (We might still have a good theory of how symbols function but, like Grice’s theory of meaning, it could not be employed in the project of naturalizing intentionality.)

Contrast this case with one where there is a clear victory for naturalization: chemistry. One of the most important chemical properties of an element is its *valence*, which was originally defined as the number of atoms of hydrogen the element could combine with (so, oxygen has a valence of 2). Lots of chemistry could at least organized, if not explained, by use of this idea. But of course the urge to naturalize arose with the question exactly what *is* valence. And it turns out that everything the old chemists were going on about with their talk of valence as a fundamental chemical property of elements can be accounted for by the physical structure of atoms, in particular in terms of the structure of the outer electron shells. This means that someone could (albeit inefficiently) learn what valence is without bothering about developing a prior acquaintance with chemistry – there is no need to understand chemistry in order to understand the physics of valence⁵. While this naturalizing story *is* a good explanation of valence, *that* it is a good explanation is not part of what makes it the proper story about valence – the world sees to that all by itself. Bio-semantics can’t let the world do its job since the world can’t distinguish between problematic contents (that’s why there is a problem about specifying content in the first place). Thus bio-semantics makes its explicit appeal to the canons of good explanation. Part of what makes ‘horse’ mean *horse* is that this *explains* why ‘horse’-symbols proliferated, and this appeal violates Rule 2.

Another well known and highly developed theory of content is that of Jerry Fodor (see ??). This theory – unlike, as I think, Millikan’s – has as its explicit aim the production of a theory that abides by The Rules. Fodor solves our *horse vs. horse or anything indistinguishable from a horse to 20th century science* problem by appeal to what he calls asymmetrical counterfactual

⁵ I am strongly inclined however to think that no one could understand the physical account of valence without already understanding what explanation is supposed to be, that is, without knowing about how minds connect information together, find certain things interesting and relevant, etc. (This does not interfere with the naturalization of chemistry since these intentional notions are not chemical notions.) I am also inclined to think that this seemingly innocuous fact is very important to the project of naturalizing the mind, and in fact probably makes it impossible (see below).

dependence. The idea is that instances of *horse or anything indistinguishable from a horse to 20th century science* would not cause a ‘horse’ unless instances of *horse* do (and did) cause ‘horse’s, but *not* vice versa. Of course, it is very hard to believe that we really do have asymmetric dependence in this case, but leave that aside (along with a host of more or less technical difficulties; see Adams+ or Seager). We are interested in Plato’s problem and it perhaps arises here when we consider whether the appeal to counterfactuals in Fodor’s theory secretly invokes some of the very notions which the theory is supposed to be naturalizing. Arguably it does, since arguably the kind of counterfactuals the theory needs do not have any determinate truth conditions independent of the *goals* of explanation (thus there is a weak affinity between Millikan and Fodor here). There might be some counterfactuals that have, so to speak, world-limited truth conditions (what Putnam called strict counterfactuals), but it is, to say the least, unlikely that

instances of *horse or anything indistinguishable from a horse to 20th century science* would not cause a ‘horse’ unless instances of *horse* do cause ‘horse’s

is an example of one of them. You can’t tell whether this counterfactual is true unless you understand the context in which it is to be evaluated, and you can’t discover this context unless you understand how counterfactuals are interest-relative, and *interest* is a thoroughly intentional notion. Another way to put this: Fodor needs to know the truth value of certain counterfactuals. Unfortunately, these counterfactuals don’t have a truth value *simpliciter* but only relative to a ‘context of evaluation’. The notion of a ‘context of evaluation’ is itself an intentional notion and to the extent that the theory appeals to it, it violates Rule 2⁶.

Notice that, as in the case of bio-semantic’s use of the notion of explanation, an appeal to counterfactuals (even non-strict ones) is no barrier to naturalization in general, since in almost all cases of naturalization the appeal to an intentional notion would not violate The Rules (since the target would not be itself intentional).

§2. Naturalism and Dennett.

Even if you accept The Rules of naturalization, there are still different ways to play the game. For example, the failure to naturalize something could be seen not as a failure of the scientific world view, but rather the *discovery* that the target was chimerical. Not only the history of science is replete with well known examples of this; it is commonly occasioned by our growth of general knowledge about the world (of course, such growth is highly conditioned by scientific progress). There is no scientific proof that demon possession is unreal (so far as I know – and

⁶ My criticism of Fodor overlooks a feature of Fodor’s theory that might provide an answer to Plato’s problem, namely his appeal to laws of nature construed as relations between properties. This is supposed to make it an ‘objective fact’ that there is the appropriate relation between ‘horse’s and *horses*. I doubt that this really makes any difference because I doubt that there are such relations independent of interest-bound constraints on the set of possible worlds we are allowed to look at (it seems clear to me, at least, that there are nomologically possible worlds not too distant from the actual world where the ‘horse’-*horse* link is quite different).

may we continue to be preserved – there has never been any attempt at a scientific demonology), but the idea has fallen by the wayside. The trend of this eliminative history reinforces the respectability of our Rule-defined notion of naturalization. It seems that where ever we have seen no prospect of naturalization we have preferred elimination to ontological expansion⁷.

Thus it is curious that Dennett, who by and large writes from a perspective that clearly endorses the scientific view of the world and which is supposed to be non-eliminativist, espouses a theory of intentionality which blocks the naturalization of the mind. Although Dennett's theory of the intentional stance is by now intricate and subtle, it remains essential to it that the mental states of a subject, S, be understood as states (no doubt physical states, probably of S's brain) whose *mentality* resides in their underpinning an intentional interpretation of S. Now, on just about anybody's view, the mental states do the job of generating behaviour which can be interpreted from the intentional stance, but for most theorists the mentalistic interpretation is parasitic upon the *mental properties* of these states. It is because they are mental that we can successfully interpret them (or their possessor) as having a mind. Fundamentally, Dennett sees things the other way around: it is because we can (perhaps, if we are to grapple successfully with the behaviour, even *must*) interpret these states (or their possessor) as mentalistic that they are mental. To take a favourite example, the internal states of a chess playing computer are *about* chess because we can interpret the machine as playing a (reasonably good) game of chess. The straightforwardness of the example notwithstanding, it would be a *deep* metaphysical error to seek for any intrinsically chess-aimed intentional states within the mechanism. An error which Dennett sees being everywhere committed in the philosophy of mind.

The problem of 'original intentionality' is thus dodged, but part of the cost of this success is the loss of naturalization (it remains open whether outright falsehood is another, and higher, cost of Dennett's views). The notions of 'interpretation', 'intentional stance', 'predictive purposes', etc. are one and all notions which generate another case of Plato's problem. This is formally obvious, but let's be clear how the problem arises. It is not that, as a matter of fact so to speak, mental state ascriptions are parasitic upon behaviour which can be interpreted mentalistically; the problem of naturalization is that we cannot *explain* what mental states *are* without appeal to notions shot through with their own mentalistic implications. You can't understand what a mind is unless you already know what a mind is, since you can't understand mentality without understanding the intentional stance, which requires you to already understand a host of essentially mentalistic concepts. Another approach to this is via the comparison of the case of the mind with that of chemistry; the two are entirely dissimilar. One can imagine learning chemistry by learning its naturalization along with a host of defined terms – at the end one would really know what chemistry was on about, although after this beginning, because of typical problems of complexity, one would have to learn to 'think chemically' to really get anywhere in chemical studies. According to Dennett, you can't do this for the mind, since you'd already have to know what a mind was to 'get' the intentional stance.

Understanding Dennett's failure (or refusal) to naturalize the mind might clarify other issues. One example is the long standing debate between Donald Davidson and Jaegwon Kim

⁷ Such a strategy has been explicitly defended with respect to the mind, first by Quine (19??) and later by Paul Churchland.

about reductionism. Davidson's *anomalous monism* rejects the naturalizability of the mind (and for reasons not altogether unlike Dennett's) but Davidson famously accepted the supervenience of the mental upon the physical, whereupon Kim presented a variety of arguments to show that supervenience entails reducibility, construed as necessary coextension of properties (see ...). But if we pay attention to The Rules we see that reducibility need not entail naturalizability. Even if for each mental property there was a physical property nomologically necessarily coextensive with it, this would not suffice for naturalization unless this coextension relation served to *explain* what mentality *was*. Somewhat curiously, naturalization is *both* weaker and stronger than reducibility as we are construing it here. For a relation much less strong than reducibility can underwrite an *explanation* of one domain in terms of another (or an explication of how the one domain supervenes upon the other), but even a necessary coextension between domains is not sufficient, all by itself, for explanation (recall the example of the height of the flagpole and the length of its shadow – though there is obviously a necessary coextension between these the latter cannot explain the former⁸). In the case of the mind, if some of the concepts needed to explain the mind are themselves mentalistic, then naturalization will be impossible, whether or not any relation of coextension holds between mental and physical properties. I thus urge that we interpret Davidson's remark that even if we discovered relations of coextension between certain mental and the physical states we would have no reason to believe it was not an accident as the claim that the discovery of the coextension would not serve to explain the mind in physical terms. Davidson's claim here can be put in old fashioned language: discovering the neural 'correlates' of mental states does not *explain* the physicality of the mind. This seems worth emphasizing since many have fallen into the trap of believing that the successful discovery of a pretty complete set of robust mental-physical correlations would amount to naturalization.

All this seems to lead to the rather nice result that the failure to naturalize the mind does not mean that the mind is non-physical, although it does entail that the mind is physically inexplicable, which is to say that the failure to naturalize does lead to a version of mysterianism.

For reasons I cannot understand, mysterianism tends to be a vilified, rather than merely criticized, doctrine, but there are several forms that deserve to be distinguished. One may hold that it is impossible to naturalize the mind because of certain inherent conceptual limitations of the human mind (see McGinn ...). Since at present we have no real understanding of the nature and creation of concepts themselves, let alone the limits of human conceptual machinery, it is hard to assess the merits of this claim. McGinn's argument depends upon quite strong assumptions about the nature of concepts and their genesis which are far from assured.

⁸ Except in very special explanatory contexts (if, for example, we imagine that someone built a pole with the intention of having a shadow of a certain length we might be able to produce a situation in which the length of the shadow did explain the height, see van Fraassen ...). Philosophy itself is a particularly clear instance of the need for an understanding of the explanatory context. Any teacher of philosophy will recognise that one of the stumbling blocks facing beginning students is that they just don't 'get it' – they don't see what philosophy is trying to do (this is usually especially evident in the inappropriate use of examples). The clearer it becomes that every explanation presupposes an *understanding* of the explanatory context, the more clear it is that the mind can never be naturalized according to The Rules.

Nonetheless, it seems plausible to suppose that there is some limit to our ability to understand the world and if so it is an empirical question whether the scientific naturalization of the mind transcends this limit.

Another form of mysterianism rests content with the neural-mental correlations, declaring them to be ‘brute facts’, incapable of explanation by *any* science aided by however powerful a conceptual system. Ironically (given his views on the first type of mysterianism), it is possible that Owen Flanagan promotes this position when he says ‘some patterns of neural activity result in phenomenological experience; other patterns do not. The story bottoms out there’ (1992, p. 58). I have some trouble understanding how a feature of the world that manifests itself (so far as we know) only when vast numbers of complex neural units interact in quite special ways can be at the ‘bottom’ of the story of the world. The bottom of this story ought to reside in the very simplest features of the world; thus the charge of the electron is perhaps a candidate for being a brute fact, but not the sensations brought about by electrical current.

There is yet another form of mysterianism, which I call *methodological mysterianism*, and which I have been urging above. It is possible that the conditions of naturalization preclude naturalization. I take this to be so – Rule 2 cannot be satisfied in the case of the mind (intentionality, meaning, content, etc.) since a variety mentalistic notions must be understood in order for any explanation to be given. This is universally a condition upon explanation but matters not at all to the project of naturalization throughout all domains that are remote from mentality (as discussed above, the naturalization of chemistry is not blocked by the fact that any explanation of chemistry in ‘purely physical’ terms assumes that there is a more or less implicit understanding of explanation, the context of explanation, the notion of intelligibility, etc.).

While I think that methodological mysterianism is a *general* impediment to the naturalization of the mind, and in fact could explain the so-called explanatory gap between matter and consciousness, and thus ease our qualms about accepting the identification of mental and physical states in despite of this unbridgeable gap, I won’t expand on that here. For there are more specific versions of methodological mysterianism and Dennett’s view of mind is one of them. Insofar as notions such as *interpretation, explanatory & predictive purposes* and the like are required to understand what minds are then Rule 2 cannot be fulfilled, but not, on the face of it, because of any ontological or conceptual problems. This is a kind of methodological mysterianism.

However, despite the fact that Dennett’s views set up only a methodological block to naturalization, one might remain unsatisfied. The mind *is* physical at bottom and there must be some account of what this amounts to, mustn’t there? Yes, and no.

§3. What *Can* be explained or Darwin to the rescue.

Anyone who read the, shall we say, heated exchange between Dennett and Stephen Jay Gould (see ...) about *Darwin’s Dangerous Idea* will have noticed, among other things, that Dennett has a strong commitment to optimality, or adaptationist, explanations in evolution. While Dennett vigorously defends adaptationist thinking with a set of argument internal to evolutionary theorizing, the importance of optimality for his philosophy stems from a deeper source.

Adaptationism is needed to save Dennett from a much stronger form of mysterianism

about the mind than the mere methodological mysterianism we ascribed to him above. Suppose that one felt the opposing pulls of an interpretationist or intentional stance theory of mind (or mental states) as well as the residual attraction of a Rule based project of naturalization. Strictly speaking, the former would preclude the latter, but there might remain something central to the favoured picture of the mind which could be plausibly naturalized. Dennett's scheme is perfectly set up for such a manoeuvre.

The nature of the intentional stance requires that there be appropriate input into the engine of interpretation, and this input is behaviour – generally speaking extremely complex behaviour (and at the higher end, mostly verbal behaviour). At least to a first approximation behaviour can be described in non-mentalistic language, although the more complex interpretable behaviour becomes the less sense can be made of it from outside the intentional stance. So one might be tempted to ask for the non-intentional description of behaviour that licences the ascription of mental states in the hope for a Rule based naturalization of mind on the basis of such non-mentalistically described behaviour. But since this strategy is simply behaviourism in a 'metaphysical' guise, it is abundantly clear that it cannot succeed. Furthermore, if it *could* succeed we would have no reason to cleave to the intentional stance as our theory of mind; we would have fully succeeded in naturalizing the mind (at least on the assumption that the physical mechanisms of behaviour generation are naturalizable, an assumption which is entirely plausible on almost everybody's view). And, even leaving aside the devastating critique behaviourism has been subject to, there is simply no hope of this, since the behaviourists would have to understand and more or less covertly deploy a host of intentional concepts (such as, to reiterate, explanatory purpose, predictive goals, relative intelligibility, etc.) in order to understand their behaviouristic theory of the mind. Thus the full naturalization of the mind would have failed, falling victim, as so many attempts do, to Plato's problem, in particular, to the version of it that supports methodological mysterianism.

In any event, there is another path. We cannot explain the mind in terms of behaviour (or behavioural disposition, behavioural patterns or whatever), yet the intentional stance still takes behaviour to be the foundation of mind; actions are interpreted *behaviour* where the behaviour at issue can be, as a matter of fact, non-mentalistically described. And it is possible to ask for a scientific account of the origin of systems which can display behaviour sufficiently rich to deserve (require?) intentional interpretation. Here we can get a lift up from Darwin.

Crudely speaking, the claim is that evolution can explain the genesis of organisms capable of ever richer patterns of behaviour including behaviour susceptible to intentional description. Once this claim is in place we seem to have an account of the origin of mind from 'mere matter' even though we lack a 'proper' Rule based naturalization of mind; call such an account a quasi-naturalization. There is an air of sleight of hand here, but this may stem from incompletely suppressed hankerings after Rule bound naturalization. Consider Dennett's account of the 'birth of meaning', which he traces back to the earliest beginnings of life on Earth whereupon arose the possibility of 'exercising the option of adopting the perspective from which errors might be discerned' (ddi, 203). What perspective, you might ask, and who is adopting it? It is we, enminded creature, that adopt the perspective but we the errors are found in the primordial replicators' failures to reproduce themselves *exactly*, and that *is* where the errors are isn't it? One could hardly chide the early replicators for failing to take the error-discerning perspective (you might as well complain that it's your golf ball's fault for going into the rough, though indeed it is

the thing sitting in the tall grass). Still there is a long way from faulty molecular transcription to folk psychology. Is there any *account* of the genesis of behaviour patterns which tend towards deserving a mentalistic interpretation?

I'm not sure that Dennett has ever ventured to demonstrate that folk psychological interpretability is a natural product of the evolution of complex organisms. From one point of view, the early part of ch. 7 of CE is close to such an attempt (see also 1978 - law of effect -). A crude story more directly aimed at integrating folk psychology and evolution might go something like this. Organisms that are more successful at replication will proliferate at the expense of the less successful. We can expect that organisms will have to compete for the resources necessary for replication (and it is this competition that will drive the creation of complexity in organisms). So the more successful replicators will out-compete their rivals. What will this amount to? It is not hard to believe that the more successful organisms will appear (at least) to make more of an effort to get the resources (this could work in a variety of ways, they might be better at finding resources, or keeping resources or taking resources away from other organisms – ancient, not always honourable, and still familiar strategies). That is to say (almost) that the successful organisms will give, within the initially stringent limits of their behavioural capacities, indications of *wanting* the resources necessary for replication. The usefulness of some method of tracking resources is clear and could (did) lead to the concomitant development of sensory organs. From the point of view of developing folk psychologically characterizable organisms, the birth of sense organs is the birth of *belief*. Since the only value of such tracking mechanisms is to ensure that the 'desired' resources are obtained, there is a natural ground for the basic pattern of rationality which grounds folk psychology: organisms will want what they need and believe what is true (albeit within a very restricted domain). Although very crude, some such story seems not implausible as a ground of at least a quasi- or pseudo psychological characterizability. We have no trouble looking even at ants from a folk psychological perspective (actually its rather hard not to look at them this way unless we make a conscious effort)⁹.

There are other clues that there is this kind of evolutionary story to be told. It seems to me striking that organisms that are extremely different by almost any measure often engage in behaviour patterns that cry out for a similar sort of (pseudo) psychological interpretation. Here is one such example. The Scottish red deer and the African funnel-web spider exhibit striking similarities in behaviour. In the rutting season, stags in possession of a harem are likely to be challenged by intruder males. Their bouts typically involve "roaring contests" (where each apparently tries to out roar the other), "parallel walks" (where the stags walk along beside each other for varying lengths of time) and, occasionally, fighting. Pretty clearly, the precursor activities aid each stag in assessing the fighting ability of the other. This information exchange and its point could easily be expressed in folk psychological terms, and in fact it is difficult not to think of these animals as engaged in a variety of cognitive tasks (and surely such higher

⁹ I have given a more detailed account of this argument in Seager (19??). There I claim that what virtually guarantees the creation of psychologically characterizable organisms is evolution's development of creatures that fulfill the theory of teleology espoused by MacKenzie, Taylor and Bennett (see ...). Once organisms can exploit what Bennett calls the 'conditional properties of the environment' they will be, more or less, psychologically characterizable.

mammals really are ‘cogitating’). Moving to another order of organism, the female funnel-web spiders contest webs, and their typical bouts involve the following: “(i) ‘locating’; orienting movements, and palpation of the web ... (ii) ‘signalling’; lengthy exchanges of vibratory and visual displays ... (iii) ‘threat’; running or lunging toward an opponent. (iv) ‘contact’ ...” (Maynard Smith (1982), 115-116). The point of the precursor activities again seems clearly to enable each spider to assess the fighting ability of the other (and possibly the value of the web¹⁰). A description of a typical encounter in terms of belief and desire would not be hard to produce. While both the style and function of these behaviour patterns in deer and spider are strikingly similar and in themselves equally susceptible to psychological characterization, it does not seem very likely that these creatures share any significant neural structure, state or process that accounts for it. The mammalian brain of the deer is immense and densely connected compared to the paltry brain of the arthropod, yet both, in the appropriate environments, produce ‘psychologically equivalent’ behaviour. This is, of course, not to say that the spider is the intellectual equal of the deer; in fact, the greater the disparity between deer and spider the better for the point I’m trying to make, which is that psychologically characterizable behaviour is a fat evolutionary target, that mother nature would have trouble missing once she started building organisms of any appreciable complexity¹¹.

So perhaps we can allow that behaviour suitable for interpretation from the intentional stance is the natural byproduct of the evolution of complexity in organisms. There remains a wide gulf between the animals and ourselves. The *human* intentional stance is incomparably more complex, subtle and intricate than that required for the interpretation of animals, as is the behaviour interpreted by the human stance (as Dennett has frequently emphasized). A problem looms here. In light of the gulf between animal and human mentality as indexed by behavioural differences, it is possible to maintain that although ‘basic’ psychological characterizability is one of evolution’s natural products, the human mind, with all its *distinctive* features, is *not* a product of selection. This would threaten even the weak naturalization of the mind which Dennett’s view can still allow, leaving the mind (the human mind at least) an accident of nature. If it should turn out that the features that distinguish human from animal mentality are *spandrels* (that is, accidental and unsought for byproducts of independent development) then we shall have no evolutionary account of the patterns of human behaviour after all. We would then be forced to say, first, that no account of the mind can be given except in terms which presuppose mind and, second, the best account of why creatures with minds like ours arose is that it was a lucky

¹⁰ But there is evidence against this in favour of the view that only the owner of the web knows its value (Maynard Smith (1982) 116).

¹¹ To be fair to the vegetable kingdom, plants are also pretty complex. The kind of complexity that matters is that stemming from a kind of ‘information laden’ interaction with the environment, especially with other organisms and mediated by sensory organs rather like our own. To some extent plants fulfill these conditions too, but it is also true that some weak psychological characterizations of plants are not foreign to us. As soon as we notice that plants are capable of certain sorts of movement, we speak of plants ‘trying’ to get to the light, or spreading roots because they ‘want’ water.

accident, perhaps, if we have still more luck, building upon the foundations of animal cognition. Though not as extreme as, for example, McGinn's, this would be a quite robust form of mysterianism: you can't understand the mind in scientific terms and it arose by a fluke of nature.

An extreme, and extremely implausible, example can illuminate the problem. Suppose that the growth of the human brain which supports our distinctive mentality was occasioned by selection pressure for blood cooling, completely independent of cognitive function. That is, we were doing just fine with our ape-brain, but the move on to the savannah forced development of better heat dissipation methods – so we shed our fur and grew our brains. As a purely accidental byproduct of this brain growth we suddenly 'woke up', became conscious, invented language and culture. According to this tale, while the complex behaviour distinctive of human intelligence emerged out of this brain growth, its accidental nature leaves the radical change evolutionarily inexplicable. Now, my story is ridiculous (there is more chance of the reverse being true, that brain growth provided better cooling as the accidental, but useful, concomitant to cognition driven growth) but the more 'accidental' the brain changes underpinning human intellect (or, equally from the intentional stance, the behaviour distinctive of human intellect) the more mysterian becomes our account of the mind.

How much of the human mind is the accidental result of neurological change occasioned by non-cognitively driven changes? Unfortunately, neither the question, nor the methods of answering it are very clear. Gould, for one, is willing to assert that 'adaptationism [is] a particularly dubious approach to human behaviour' since

Many, if not most, universal behaviours are probably spandrels, often co-opted later in human history for important secondary functions. The human brain is the most complicated device for reasoning and calculating, and for expressing emotion, ever evolved on earth. Natural selection made the human brain big, but most of our mental properties and potentials may be spandrels – that is, nonadaptive side consequences of building a device with such structural complexity. (From 'Darwinian Fundamentalism' ...)

On the other hand, if we are allowed to use *language* as the index of achievement of the human mind, there seems to be evidence of structural change in the brain aimed at supporting linguistic functions (see Deacon ...). Could language be a spandrel? Here is one place where the unclarity mentioned above intrudes. As Dennett points out, from a certain point of view, all – certainly most – of the features of highly evolved organisms probably began as spandrels, but if one has to go back a very long way to spot the spandrel nature of the feature then there is an intervening adaptationist story of the development of the completed feature from the initial 'spandrel' which certainly *looks* like it provides an evolutionary explanation of that feature. Furthermore, even if language developed from some neurological feature unrelated to linguistic function in some more interesting sense (for example, if the supporting feature is very recent), what grounds are there for thinking that this unrelated function was non-cognitive? Both Deacon (19??) and Donald (19??), for example, suggest that a pre-linguistic, but *symbolic* function spurred the unusual development of the frontal cortex which now supports language. Perhaps there is an evolutionary story of the development of abstract symbolic abilities that is analogous to the story of the development of sensory organs. If so, then there is some possibility of extending the quasi-naturalist account of elementary folk psychology to the level of the human mind. I take it that one of the basic projects of 'evolutionary psychology' is, or should be, to bridge this gap. And while

it has its detractors, it is too early to tell whether evolutionary psychology will fail at this task (though it is a much harder job than, for example, explaining why more men than women are philanderers or why men murder more often than women – the sort of thing most evolutionary psychology seems to spend most of its time on).

One might think that Dennett has another option here, which is to run the quasi-naturalization at a different level. Perhaps much of what is distinctive about the human mind has been produced by the human mind, via the development of *memes*, a process which is itself susceptible to adaptationist explanation. But there are two problems here. In the first place, it is unlikely that the analogy between genetics and natural selection on the one hand and the propagation and development of memes between and within human minds on the other is more than superficial. Genes are nicely atomistic and discretely compositional; ideas are not. So a memetics based on the model of genetics probably won't go very far. The second problem is worse, at least from the naturalist's perspective. Memes are products and inhabitants of mind *as such*. We have no understanding of them except as units of meaning. Thus, if we plug them into the quasi-naturalist story we are trying to develop we once again fall victim to Plato's problem. The previous project avoided this by looking for an account of the behaviour patterns that sustained interpretation from the intentional stance. This behaviour was supposed to be characterizable in non-mentalistic terms and we were supposed to be able to see how natural selection could mould behaviour thus characterized into appropriate – mentalistically interpretable – forms (recall that, even on the most optimistic assessment of this project, *full* naturalization fails since we always presuppose the intentional stance itself). This is not to say that the meme-story is not worth telling or somehow illegitimate, but that it cannot replace the original, evolution based, quasi-naturalist program.

That program seems to me coherent, fruitful and integrated in a variety of interesting ways with a host of 'cognitive sciences' (biology and neuroscience as well as AI and psychology). To summarize: it takes as given the intentional stance, that is, our core mentalistic notions by which we interpret the behaviour of our fellows (and ourselves). Thus it must forsake Rule bound naturalization. The resulting mysterianism is, I argued above, a kind of methodological mysterianism which is general and unavoidable. It leaves behind an outstanding debt to the scientific world-view which can be only partially repaid by an evolutionary account of the genesis of the behaviour patterns which are the targets of the intentional stance. To the degree to which this promised account lapses into appeals to accidents or spandrels to that degree the underpinnings of mind remain mysterious¹². Notice that there are two *kinds* of mystery here. One

¹² It is of course possible to believe that the physical properties of the brain generate all of our behaviour. Everyone *does* believe this already. The problem is to extract from the neural details an *explanation* of our behavioural capacities, and specifically those behavioural capacities that underwrite the ascriptions of mind of the intentional stance as applied to complex, distinctively human behaviour. I suspect that the gap is too large here, and the project would be akin to *understanding* plate tectonics in terms of quantum chromodynamics. The beauty of the adaptationist story is (or would be) that it tells us why the neural details *have* to generate psychologically characterizable behavioural capacities. Then we can happily dig away explaining how neurons account for little bits and pieces of these capacities.

is the methodological mysterianism which stems from the fact that there is no way to explain the mind except in terms which essentially depend upon other mentalistic concepts. The second is the more familiar sort wherein an evolutionary account of the genesis of certain behavioural capacities may be difficult or impossible to come by. We can hope to overcome the second sort of mystery, but never the first. It is my hope that methodological mysterianism is entirely safe for human consumption (and might be a pleasant and relaxing brew). I want to conclude however with my fear that it is not, which leads to a strange and interesting confluence of views.

§4. Patterns and metaphysics.

The question, at bottom, is whether methodological mysterianism has any impact upon our metaphysics. I mean by metaphysics simply our general picture of the world, our picture of how things ‘hang together’ in the large. The picture behind The Rules of naturalization is such a metaphysics and one that is very powerful and compelling – call it the scientific picture of the world (or SPW for short). What is the SPW? Here is a sketch (which should not be unfamiliar).

To begin at the Beginning, we have a pretty good idea of how our universe began, though we await the long anticipated marriage of quantum mechanics with general relativity for the *really* early details of creation. The theory of the Big Bang reveals how matter originated and suggests how early inhomogeneities in energy distribution led to star and galaxy formation. We know how stellar processes create the heavier elements and how massive dying stars ‘fertilize’ space with these newly generated elements. Such second-generation material forms new stars and planetary systems some of which are lucky enough to have planets orbiting neither too far nor too close to their sun, and possess the chemical mix required for life.

While no one knows how life originated on Earth (let alone anywhere else), this is a scientific question. Those toiling on the creation of a scientific metaphysics, have no reason to deny that life is just a matter of complex chemistry. From, as it were, the bottom up, chemistry itself is understood as the natural outcome of the physical interactions of the atoms’ various constituents. Life is the pinnacle of chemical complexity.

Within the realm of living things the hierarchy of chemical complexity is again extended, from very simple structures not fully alive to exceedingly complex creatures, some of whom aspire to produce science and the SPW. What is more, we possess an exceptionally elegant theory of how biological complexity arises. Again the SPW integrates the top-down and bottom-up perspective. From the top, the theory of evolution tells us how organisms become more complex, so long as complexity is reproductively advantageous, which it evidently was within many of the environments the earth has provided over the last three billion years. From the bottom, we have forged an indissoluble connection between chemistry and genetics which has revealed the chemical basis of life and evolution.

Take anything you like: a galaxy, a person, a flounder, an atom, an economy – it seems that anything can be *resolved* into the fundamental physical constituents, processes and events which determine its activity¹³. Although innumerable difficult questions arise at every stage of

¹³ It may be worth noting here that this is not an endorsement of so-called ‘part-whole reductionism’. We know from quantum mechanics that the states of ‘wholes’ are not simple

resolution, and there is no practical prospect of *knowing* the full details of the physical resolution of anything much more complex than even a simple atom, the picture is clear. And since the *world* has no need to know the details but just runs along *because* the details are the way they are, the problems *we* have understanding complex systems in terms of fundamental physics are quite irrelevant to the metaphysics of the SPW.

Here's a thought experiment to assess one's attitude to the SPW. Imagine a computer simulation of a part of the world. First restrict attention to something 'simple' – a pendulum swinging on the moon say. The simulation covers a restricted region of space and time (though 'boundary conditions' can be set up to represent external influence), and must be defined solely in terms of fundamental physical attributes. The programmer is *not* permitted gross parameters such as the mass or the length of the pendulum, or the lunar gravitational force, but must write her code in terms of the really basic physics. (It might help to imagine the code written in terms of the properties of the atoms of the pendulum, its support structure and moon, though these are not really physically basic¹⁴.) The SPW predicts that the simulation – appropriately displayed – would reveal a pendulum swinging above the lunar surface. Now up the ante. Imagine simulating a more complex situation, for example a child's birthday party. Do you think the simulation would mimic the actual party? Though there is, of course, no prospect of ever being able to develop such simulations, I think the notion is perfectly well defined (or could be with a little more effort). I venture to maintain that what we know about the world strongly suggests that such a simulation would 're-generate' both the action of the pendulum and the behaviour of the children. I think that one way of understanding the task of physics is exactly as the development of a theory of the fundamental features of the world whose simulation would meet this challenge.

functions of the states of their parts but this does not tell against the characterization given in the text. Quantum mechanics is a *celebration* of how the interactions of things can be understood – rigorously understood – to yield new features. It is, if you like, the mathematical theory of emergence.

¹⁴ Real versions of something like my imaginary scenario now exist and are already fruitful. For example, there are computer models of quantum chromodynamics that can compute the theoretically predicted masses of various quark constituted sub-atomic particles (see Weingarten 1996). The ultimately insuperable problem of computational intractability is all too evident, for realizing these calculations required the development of special mathematical techniques, the assembling of a dedicated, parallel supercomputer specially designed for the necessary sorts of calculations (a computer capable of 11 billion arithmetical operations per second) and roughly a year of continuous computing. Weingarten reports that a special 2-year calculation revealed the existence of a previously unrecognized particle, whose existence could be verified by examining past records from particle accelerator experiments. Modelling the interactions of particles would be a much more challenging task, suggesting to the imagination computational projects analogous to the construction of medieval cathedrals, involving thousands of workers over many decades.

In any event, if you think the simulation would agree with reality, you believe in the SPW¹⁵.

The problem I'm worrying about emerges upon consideration of the status, within the SPW, of the myriad of high-level explanatory structures ubiquitous throughout science and ordinary life, of which the primary example is the complex behaviour interpreted by the intentional stance. These Dennett calls patterns.

Inhabiting a curious zone midway between, as it were, objectivity and subjectivity, patterns are *there* to be seen, but have *no function* if they are not seen. By the former, I mean that patterns are not just in the eye of the beholder; they are really in the world and provide us with an indispensable and powerful explanatory and predictive grip upon the world. By the latter, I mean that the *only* role they have in the world is to help organize the experience of those conscious beings who invent them and then think in terms of them. That is, although the world is rightly described as exemplifying a host of patterns, the world has no use for them. In terms of our thought experiment, high-level patterns do not need to be coded into the world-simulation in order to ensure the accuracy of the simulation, and this is just because it is the fundamental features of the world which organize the world into all the patterns it exemplifies and they do this all by themselves, with no help from 'top-down' causation. Doubtless there is a harmless sense of 'top-down causation' which is perfectly acceptable, appropriate for use within pattern-bound explanations. For example, we can explain the location of a particular atom by reference to the *intentions* of the operator of a scanning tunnelling electron microscope. But we know that those very intentions are elusively accommodated within a vastly intricate web of micro-states which, within their environment, 'push' the target atom to its final location. Intentions, like planets, animals and molecules, have no need to be specially written into the code of the world-simulation.

We could define 'radical emergence' as the hypothesis that the world-simulation will fail, save for the addition of certain high-level, complexity induced properties ready to step in and 'make a difference' at various critical junctures. The SPW, then, is the denial of radical emergence. The SPW says that the world generates complexity out of simplicity, but never dreams that complexity itself has powers that outrun its source. There is another notion that is sometimes confused with radical emergence but is quite distinct which I'll call 'explanatory emergence'. This is the doctrine that complexity does outrun the *explanatory resources* provided by an *understanding* of the simple. Explanatory emergence is compatible with radical emergence but obviously does not entail it.

Now we can see that The Rules of naturalization express the hope that explanatory emergence is false, or, rather, that it is not wholly true; that there is an *explanation* of every high-level phenomenon in at least slightly lower-level terms (and thus by a chain of explanations the world will be unified in accordance with the ultimate primacy of its fundamental physical

¹⁵ Some would be inclined to say that we already have one computer running the perfect simulation: the world itself (see Deutsch ...). This is another way to test your attitude. If you think this idea must be right, that the world, at least, is a powerful enough 'simulator' and that the 'machine language' it's running on is basic physics, then you believe in the SPW.

features)¹⁶. By and large, this hope has been fulfilled. Explanation is a less stringent formal requirement than outright reduction¹⁷, and the SPW is chock full of Rule based explications of high-level phenomena, all of them tending to follow explanatory pathways back towards the truly fundamental features described in physics. The explanations which Rule based naturalization require are the sort that let us understand what a phenomenon *is* in lower level terms, but they are not required to be so strong as to replace high-level discourse. We do, I think, understand what chemistry is in terms of physics, but it is not possible, and there is no desire, to replace chemical talk with purely physical talk. In terms of our world-simulation thought experiment, we can express all this very simply – we would expect to see chemistry (or simulated chemistry) at work in the simulation even though no chemistry was coded into the simulation. Our belief in this feature of the imaginary simulation is an expression of our belief that there is no radical chemical emergence. And while perhaps we don't know *for sure* that radical emergence is false, that's where the smart money has gone for the last 400 years.

But we've seen that Rule based naturalization can fail in ways more insidious than radical emergence. Because of Plato's problem, the theory that mind is to be properly understood in terms of interpretation from the intentional stance precludes Rule bound naturalization. Dennett's ideas about patterns involves a generalization of the interpretationist scheme, for patterns are explanatory posits, useful only from a point of view which understands them (bearing in mind that understanding comes in degrees).

Perhaps this is disputable. Isn't the world full of patterns in use by a host of systems to which we would be unlikely to ascribe mind? For example, don't some moths have eye-shaped patches on their wings the function of which is to startle predators and thus giving the moth an extra fraction of a second to escape being eaten? Doesn't this work – doesn't it *have* to work – via the predator's *recognition* of the patches as eyes? Well, is the moth *trying* to scare the predators when she flashes her wings? Now, of course, at some point mind does intrude and we do have recognition and intentional deception. And there we have the patterns taken up and understood by the only thing that can understand, appreciate or perceive patterns – minds. Precisely where the changeover from non-mind to mind occurs is a vexed question (even allowing that the distinction will be fuzzy), but what matters is the point that *patterns* have no role to play in the world unless and until they are taken up in understanding by minds.

¹⁶ Bear in mind that 'x explains y' is *not* a transitive relation. Nonetheless, a chain of linking explanations suffices for an appreciation of the world's unity; our slogan could be 'the world is one, science is not'.

¹⁷ We might define 'reductive emergence' as the claim that there are high-level phenomena that cannot be strictly reduced to low-level phenomena, where reduction is conceived in classical terms (i.e. the high level laws are to be strictly derived from low level theory plus the 'bridge laws' which provide extensional definitions of the high level phenomena at issue). I think most would accept reductive emergence which entails the explanatory indispensability of high level theory. The combination of reductive emergence plus the denial of radical emergence might be called 'benign emergence' which is, I think, the generally accepted position today. But I stress that reductive emergence does not entail explanatory emergence.

Does evolution work by discovering, creating and arranging patterns? We understand evolution as a ‘search’ for optimality, but the search is blind. Any intentionality ascribed to nature is entirely and only metaphorical, the appreciation of which is dependent upon a prior understanding of mind. The patterns we see in evolution are there, but are there only to be *seen*. Nothing but mind can see them and only mind ‘needs’ them. Evolutionary theory is a way of apprehending parts of the world which brilliantly highlights certain more or less enduring patterns to be found in the ‘dance of the atoms’. *We* need the idea of evolution to understand the world, but the *world* has no need of it; the world – not even the biological parts of it – is not being driven by evolutionary properties. Predator-prey relations will be revealed in our imagined world-simulation of an environment in precisely the form they take in the world itself; the Hardy-Weinberg law will emerge from the quark/lepton/boson sea of our simulation with not a jot of evolutionary theory needed in the underlying program (unless of course, and perish the thought, evolutionary properties are more than just benignly emergent).

Leave aside the metaphors of ‘mother nature’ and it is clear that we have at least the good beginning of a Rule based naturalization of evolution, and while this will never eliminate our need to understand the world in evolutionary terms, it does reveal to us what evolution is in physical terms.

I’ve argued that is not the existence of patterns but the *function* of them which is mind-dependent. Perhaps this is made clearer by focussing on an interesting feature of some patterns – they can exist and retain their explanatory usefulness even when their naturalization reveals them to be *less* than we thought they were (here we see the source of the curious attraction / repulsion that Dennett’s views have for eliminativism about the mind).

Consider, for instance, the Coriolis force, which gunnery officers have long had to take into account when computing the trajectory of long-range cannon shells. (A host of other activities ‘require’ cognizance of the Coriolis force as well.) This is a benignly emergent property of the earth, or any other rotating system. But there is no such force; it is an artifact of a certain viewpoint. At least, if we really thought there was such a force, with its own causal efficacy, the world would end up being a much stranger place than we had imagined. Just think of it: rotate a system and a brand new force magically appears out of nowhere, stop the rotation and the force instantly disappears. That is radical emergence with a vengeance. Luckily, there is no need to posit such a force. The Coriolis phenomena are related to the underlying physical processes in a reasonably simple way – in fact simple enough for us to comprehend quite ‘directly’. We can give a perfect Rule based naturalization of the ‘force’ which reveals it to be non-existent! But the pattern retains its usefulness and no one faced with the problem of understanding how objects move over the Earth is going to give it up. And, of course, the world-simulation automatically duplicates Coriolis force behaviour patterns from the bottom up.

The final problem to address here emerges from the mind-dependence of the function of patterns: since patterns only function to help minds organize their experience of the world, they would appear to be metaphysically otiose. We have the fundamental physical structure of the world, and, according to the SPW, this is quite sufficient all by itself to generate all the patterns that mind can appreciate. Furthermore, the vast majority of patterns seem to be susceptible to Rule based naturalization, some of these approaching old-time strict reductions, others allowing only a chain of less strict explanations which reveal the principles by which the underlying physics generates the patterns at issue. But whether or to what degree we believe in explanatory

emergence is an epistemological question. The *metaphysics* of the SPW is clear and austere. Helping itself to all of modern science, it is a comprehensive and grand view of the world incorporating every day new and exciting science linking everything from quarks to galaxies. It would, in fact, be perfect if it weren't for mind itself (consciousness really is the heart of the problem for patterns are required only insofar as there is conscious apprehension of them – the SPW does a perfectly good job on unconscious apprehension).

The more general view of how patterns function in explanation (of which the intentional stance is but one example – we see now that there is, in effect, the 'chemical stance', the 'biological stance', the 'Freudian stance' etc.) reveals that the methodological mysterianism we encountered above may, after all, be in conflict with the SPW. Methodological mysterianism tells us directly that the mind is an explanatorily emergent phenomenon. Very well, we say, but mind nonetheless *is* a physical phenomenon at bottom. Unfortunately we cannot rest there. Mind cannot be 'just another' pattern. If it was, it would be metaphysically otiose. But it cannot be, since the role of patterns within the world depends upon minds' appreciation of them. This seems to be another version of Plato's problem: we cannot explain mind as pattern since pattern depends upon mind. The SPW likes to 'dissolve' patterns into the swirling micro-machinery of the world, but the application of this strategy to the mind yields not an account of the place of mind in the world but rather an eviction of mind from the world. But the SPW actually *requires* minds as a part of its overall picture of the world. *Other* patterns can be integrated into the SPW as structures noticeable to minds. Given minds, this strategy works well and smooths over the few disagreeable aspects of the SPW's otherwise very attractive picture of the world summed up by our simulation thought experiment. But minds themselves cannot be similarly integrated without falling into a vicious circularity.

It would put too much stress on the SPW to admit that the mind was, in its terms, utterly inexplicable or completely outside of its purview. Perhaps one could fall back to the position that the link between the basic physical features of the world and mind was a 'brute fact', but this is unattractive for at least two reasons. One we discussed above, the notion that there are 'brute emergences' linked to the formation of certainly highly complex material structures is entirely unappealing and utterly at odds with the outlook of the SPW (in fact, the metaphysics of patterns is the replacement of the idea of emergence as a feature of the world and seems altogether superior an account). A second difficulty is that such a brute emergent must be entirely inefficacious insofar as the SPW rejects, as I think it must, radical emergence.

I can only suggest that the solution is perhaps to recognise that the metaphysical picture of the world that accords with the pattern based view of explanation is in the end radically unlike the SPW. I think one philosopher who draws a version of this conclusion is John Haugeland (see ...), but the effort I've expended in drawing this out so long stems from a general sense I have from all of Dennett's writings that the SPW is rather attractive¹⁸. What might an alternative

¹⁸ I note that in 'Real Patterns' Dennett gestures admiringly towards Arthur Fine's Natural Ontological Attitude as an appropriate version of scientific realism. The interpretation of NOA is far from clear but it would seem to support the SPW as outlined. Crudely speaking NOA asserts that if, for example, you wanted to know if electrons exist you should go and ask a physicist (rather than a metaphysicist, i.e. a philosopher). NOA is an attempt to banish metaphysics from

metaphysics, more in keeping with the pattern based theory of explanation, look like?

I call it surface metaphysics (SM is, I'm afraid, an all too appropriate abbreviation for it). It begins with the claim that our basic physics is no less a 'pattern' (or an appreciation of certain patterns) than any other theory. And it assumes that patterns are built out of the experiences of conscious 'observers'. Since the patterns that 'correspond' to mature scientific theories are very remote from experience, SM allows that patterns are built up recursively, out of preexisting patterns (or such patterns plus experiences).¹⁹ We must also allow that minds are themselves moulded by the patterns they take up, so that experiences are not 'pattern-less' but are more or less infected with the patterns passed down to us through the ages, both from earlier human efforts and even the efforts of our (enminded) ancestors (these patterns are, of course, the memes, content infected right from the start).

Even at such an early stage of development, SM is recognisably empiricist, but it is a modern empiricism incorporating the so-called 'theory ladenness' of observation. The attitude towards science which SM ought to espouse has, I think, already been developed at length and with great elegance, though perhaps with not such persuasiveness, by Bas van Fraassen (see ...) in a view he calls *constructive empiricism*. On such a view, science is in the business of generating models that help us manage the experienced world, and is no route to the metaphysical, or even – surprisingly, but there is no escape from the conclusion – the physical, depths. It seems that there may be a deep connection between constructive empiricism and the mind-body problem, creating the interesting possibility that the problem of consciousness will provide additional support for the kind of scientific anti-realism (or 'arealism') constructive empiricism enjoins. So far as I know the debates about van Fraassen's work have not explored this territory. The contortions – ever more convoluted – which philosophers have gone through to find an acceptable link between matter and mind may be a sign that a common assumption is being dangerously over-stressed. Perhaps that assumption is scientific realism.

Thus I'm urging that Dennett ought to take consciousness and the experienced world as the foundation of existence and forthrightly dismiss the demand for an explanation of *them* in scientific or quasi-scientific terms. He ought to regard science as explicable in terms of mind rather than mind in terms of science. This would be in line with the methodological mysterianism inherent in the interpretationist theory of mind based upon the intentional stance as well as the more general pattern-based theory of explanation. Dennett does seem to be willing to take

philosophy of science. Whether or not that is a good idea, it is obviously not a good way to draw out a metaphysical picture of the world which is based upon science (and if NOA firmly rejects this project then it will end up being not so far from the arealist position I am advocating). Similarly, if you wanted to know if the total behaviour of the world was generated from the fundamental physical structures you should go and ask a physicist – and you know what the answer would be. It seems quite clear that a basic goal of physics is to discover *the* physics which underwrites the SPW and the physicists have had immense success at this.

¹⁹ To be a little more precise: a pattern is defined as (1) any salient set of experiences, (2) any salient set of patterns, (3) any salient set of patterns plus experiences. Of course it would be hard to define 'salience' here.

‘intentional interpretation’ as a primitive element of his theory, and this notion is of course a mentalistic one. His theory of consciousness is dependent upon this notion and so is infected with mentalistic concepts.

At the same time, SM does not deny the indispensable importance of scientific models and goes so far as to endorse a theory of explanation that re-legitimizes all of our scientific explanations (essentially – maybe distastefully – by cutting explanation off from *truth*, here the echoes of Dennett’s own ‘instrumentalism’ about mentality are clear²⁰). We can have our science, but we can’t have the SPW. This would be the end of a old dream, the dream that energised the whole scientific revolution, but a view of the world that takes mind or ‘patterns’ (which are dependent upon mentalistic concepts for their understanding) as basic may require no less.

William Seager
University of Toronto at Scarborough

²⁰ It is interesting that van Fraassen sees his ‘instrumentalist’ (officially constructive empiricist) approach to science growing from the bottom up (quantum mechanics is a haven for scientific anti-realism). Perhaps instrumentalism can also grow from the top-down from an argument based upon a pattern based notion of explanation. (For more on how the explanatory power of experimental science can be integrated with a non-realist outlook see my 199?).