THE NATURE OF NATURALISM

Graham Macdonald and Philip Pettit

I—Graham Macdonald

I

Two varieties of naturalism have been prevalent in 20th Century philosophy: one stems from a pragmatist concern with the primacy of practice over theory, the other from a scientific interest which promotes theory to the front, often at the expense of common practice. In this paper an attempt will be made toward reconciliation. The two, it will be claimed, can be united without serious damage being done to the fundamental concerns of each. More attention will be paid to scientific naturalism, as many of the stumbling blocks on the path of reconciliation are placed there by misguided adherents of this approach. The strategy will be to show that the best way of satisfying the interests of the scientific naturalist will lead to a position which a modified Wittgensteinian could accept. At the end the rejection of theory by the Wittgensteinian naturalist will be confronted.

It has been said that any naturalistic theory must be informed by two general constraints: all truths must supervene upon descriptive truths, and the explanations on offer must be a species of causal explanation. The supervenience constraint will restrict the class of modal and normative truths allowed, as ‘bizarre’ constructions of causality are inconsistent with the naturalist’s desire to rule out unnatural, or supernatural, causal interaction. However, it is a moot question as to whether all causal truths supervene on non-causal, descriptive truths, so perhaps one should replace descriptive-truth supervenience by the nearly empty requirement that all causal

1 These are constraints mentioned by Christopher Peacocke in ‘Content and Norms in a Natural World’, in E. Villanueva, ed., Information, Semantics & Epistemology, Blackwell, Oxford, 1990. Peacocke makes it clear that he thinks the restriction to causal explanation still allows for a wide variety of such explanations.
explanations must invoke only natural causal interaction. One gets some specificity by saying what counts as natural, and for many naturalists this specificity emerges from an ontological preference for physical worlds. The only type of causal interaction which is deemed natural is physical causal interaction.

This emphasis on physical causation may not be suitable for all naturalist projects. The attempt to naturalise morality, for instance, will concentrate on its supposedly unnatural normative commitments rather than on problems with causal interaction, and here the stipulation that the normative supervene on the descriptive will be more appropriate. My interest here, though, is in that species of potentially revisionary naturalism which appears to conflict with Wittgensteinian quietist naturalism, and it is my claim that the former acquires at least some of its revisionary charge from its understanding of issues regarding causal interaction. This naturalism, after all, has a scientific bent, and it is worth asking why anybody might deem it necessary to absorb a subject matter, a domain of properties, into science. For some naturalists the answer is self-evident: to be real a property must fit into some scientific account of the world. An even stricter requirement would be one that recognised as real only nomological properties. These requirements, I will assert, are unnecessarily strict. A more plausible answer will cite the causal efficacy of those properties, and then go on to argue that the only way of making natural sense of such causal power is by showing how it is accommodated by causal nomologicality. Consistent with this view of what makes for an efficacious causal property will be an account of the explanations provided by the relevant properties: it will be claimed that the explanations must be essentially general in order to be genuinely explanatory, and a system of general explanations which cite causally efficacious properties is just what a scientific theory is.

It may be thought that the explanatory debt acquired by the acknowledgement of causal power can be paid off by crediting the contested subject area with the capacity for singular causal explanation. However, the Hempelian tradition would have us

---

2 This was the thought in Cynthia and Graham Macdonald, ‘Mental Causes and Explanation of Action’, Philosophical Quarterly 36 (1986), pp.145–58.
believe that such singular explanations are essentially incomplete, so that the explanatoriness of such explanations depends upon their implicit generality. The finding of appropriate projectible generalisations is seen as the task of science, with the threat being that without the discovery of laws we do not have genuine explanations. This drive to theoretical explanation will also be fuelled by the afore-mentioned beliefs (often only implicit) about the nature of causality: that it is inherently general, so that causal interactions must be nomologically underwritable.

It is quite difficult to escape from this demand for explanatory generality. Some of those who champion the completeness of singular causal explanations do so by sleight of hand, claiming that the properties cited in the singular explanation give one all the generality one needs. Take the following defence of the completeness of at least some singular explanations:

When o's being F fully explains o's being G—it is o's being F that explains o's being G. Explanatory impact is being carried by properties and there is generality built into the singular explanation by the properties themselves, without the inclusion of a law. This implicit generality surely implies that other relevantly similar Gs which are F will get the same explanation that o got.3

The thought seems to be that the implicit generality carried by the properties is required for the singular explanation to be explanatory, a thought which renders its singularity suspect. Take singular causal explanation. Hempel was (usually) clear that what was being explained was a particular event, in our terms the instantiation of a property, say G₁. A singular explanation of this event could cite another property instance, the instance which caused G₁, say F₁. At this stage there is no generality, even though we are mentioning properties in the explanation. Other instances of F may not have any tendency to produce instances of G. In order to get generality we would need to be told that the properties are causally stable, that their causal potency does not vary from instance to instance, and this is all that the Hempelian needs. Providing this information is the work of the nomological part of the deductive-nomological

concept of explanation, and without this information the singular explanation will be deemed incomplete.

A more radical singularist may insist that the singular explanation without information about causal stability is a complete explanation. After all, what it does is tell us what caused $G^i$, and what more does one need in order to explain that event? Who cares about what happens with other instances of $F$ and $G$ at other places and times? Whilst sympathetic to this I think more must be said to avoid the drive to theory. The radical singularist may come in two stripes, modest or immodest. A modest singularism asserts explanatory completeness but admits that generality is indirectly forthcoming *via* beliefs about the essentially general nature of causality. For reasons *independent of any requirements on explanation*, the modest singularist thinks that singular causal claims will entail causal nomologicality. In this case the naturalist will still think that theoretical work is there to be done. The immodest singularist embraces causal anarchy; there need be nothing further entailed by the claim that $F^i$ caused $G^i$. Even this minimal concession to explanatory demands is sufficient to generate further problems: singular mental causes, say, have the troubling capacity to produce physical effects, and this needs to be explained. (The restriction on causal interaction to physical causal interaction is designed precisely to placate such worries). For present purposes we will ignore the immodest radical singularist, if only for the reason that most 'scientific' naturalists are not of this persuasion, and our strategy is to provide a rationale from within this naturalist camp for moving towards a Wittgensteinian naturalism.

The scientific naturalist, then, is one who believes that there is a putative domain of related properties whose existence is threatened unless science can give some account of the causal efficacy of those properties. Unfortunately the constraint on causation imposed by the scientific naturalist, that all causal interaction be physical causal interaction, is not as simple as it appears: there is a reading of the requirement that all causality be physical which is manifestly too strong. That reading says that the only properties which are causal

---

4 James Woodward is such a modest singularist. See his 'Are Singular Causal Explanations Implicit Covering Law Explanations?', *Canadian Journal of Philosophy*, 16, 1986.
properties are physical properties. It may be surprising to naturalists, but this is not true. Or, to put it less bluntly, if naturalists insist on only physical property causality, then they will have to eliminate biology from their permitted sciences, as biological-functional properties are not physical properties, but are properly thought to have causal powers. There is a weaker reading which is in keeping with the naturalist project, and that is to make all causal powers determined by (supervene on) physical causal powers. For some, this supervenience claim alone produces the pressure for the supervening causal properties to be part of a nomologically based, causally explanatory science. And conversely, there are those who think that because the non-basic sciences do not traffic in nomologically based causal explanations, the supervening properties are not causally active. The first view takes it to be obvious that (at least some) supervenient properties are causally active, and takes it as a necessary condition on its causal efficacy that the property is one in virtue of which one event is nomologically sufficient for another event. (The second position simply reverses this line of reasoning. It assumes that the supervening properties are not nomologically projectible and concludes that they are not causally powerful.) Intentional properties are causally powerful, so there must be an intentional science. The causality located in supervening properties will be non-basic causality, mediated by lower level mechanisms. As a consequence, the laws of any supervening science will be ceteris paribus laws. The satisfaction of the antecedent only produces the effect if the right conditions obtain, and we can never know exactly what the ‘right’ conditions are.

5 For functional properties to be physical properties a reduction of biology to physics would be required; that there is no such possibility is argued in Graham Macdonald, ‘Reductionism and Evolutionary Biology’, forthcoming in Kathleen Lennon and David Charles, ed., Explanation, Reduction, and Realism, Oxford University Press, Oxford, 1992. The bio-naturalism of Ruth Garrett-Millikan makes a virtue out of the non-reducibility of biological-functionality to physical causality. See her Language, Thought, and Other Biological Categories, M.I.T. Press, Mass., 1984.


Now there is something right about the idea that the causal shape of the supervening property must be different from that of the base property, and that it (the causal profile) is in some way incomplete. However, this way of putting these points is too bare; it leaves too many unanswered questions. Leaving aside the questionable claim that one can have causal nomologicality combined with a free-floating ceteris paribus clause, the causal picture we are left with remains puzzling. How are supervening and base property related to each other? And how are they related to the effects they bring about? Won’t there be overdetermination whenever a supervening property is causally efficacious? These are large problems which will not be fully dealt with here, partly because they have been addressed elsewhere, but mainly because I wish to concentrate on the second constraint on naturalism. So section II briefly outlines a solution to the problems arising from supervenient causation, section III criticises alternative proposals, with section IV addressing some of the substantive, potentially revisionary, claims made by some naturalists which arise from the considerations about causality. Section V returns briefly to the reconciliation of the two naturalisms.

II

Our supervenience claim seems to provoke the following dilemma: either both supervening and base causal properties are implicated in the production of an effect, in which case a surprising overdetermination results, or else they are not, in which case we seem driven to accept epiphenomenalism for the supervening properties. Both epiphenomenalism and massive co-incidental overdetermination are unacceptable. (The overdetermination is accidental because there is no explanation of why the two separate causal properties should produce the right effect, why they should be so harmonised.) Elsewhere the following position has been


9 See the papers cited in footnotes 2, 5, and 8.
defended: the causal efficacy of supervening properties is ensured by such properties being co-instantiated with physical properties. On this view it is only one instance which is causally efficacious, thus ruling out overdetermination. The instance is an instance of both a physical and supervening property (a mental property, say,) thus ruling out epiphenomenalism.

The only puzzle remaining is why such a neat solution is not more widely accepted. Two reasons come to mind. The first queries the co-instantiation of supervening and base (physical) instances. How does this come about in a way which removes the coincidental character of the overdetermination account? The second is a more general worry regarding the locus of causal power; the happy coincidence of two properties does not, it is claimed, make both equally causally efficacious. The supervening property is said to be at best only 'derivatively' causally efficacious. Or else it is said that the instance is not causally efficacious qua the supervening property. Both worries seem misguided. Take the first. One example is provided by computer programmes: we design computers and programmes so that the expected result will be achieved. This shows how the instantiation of a higher order property can ensure the instantiation of a relevant lower order property: the computers are designed so that the instantiation of the program properties ensure that the lower order (hardware) properties produce effects which realise the intended higher order properties.

A problem with this example is that computers are artifacts, and what we need, so it is averred, is a naturalistic solution. One is provided by evolutionary biology. We now feel confident that there is a solid, naturalistic explanation for the design we see in nature: that design is due to natural selection. Certain properties are selected (and so proliferate relative to certain other properties) because of the effects which their instances have. In biology, these effects are, ultimately, those which enhance the reproductive

10 See Cynthia and Graham Macdonald, 'Mental Causes and Explanation of Action'.
capacity of the items possessing the properties. Given genetic transmission of these properties, they will spread in the population, such proliferation being a consequence of those effects. Once the properties have been thus ‘chosen’ to produce their effects we can say that it is the biological function of these causes to produce the effects for which they were selected.

An example will show how this happens. Amino acids are composed of triplets of base molecules, the four bases being symbolised as A, C, G, and U. Given that each amino acid is composed of only three of these bases and that there are only twenty amino acids, it is obvious that there are redundant triplets. It turns out that there are different bases which encode for the same amino acid; alanine is composed of either GCU or GCC. The change in the base does not change the effect; the amino acid sequence of the protein is not changed. The function of alanine (the effect for which it was selected) is unchanged by the change in the base. It is because the advantageous effect remains the same that one can say that the different bases share the same functional property. In the example given, the difference in the base properties does not look large, so it may be thought that larger differences must make a functional difference; but this is just wrong. The way functional explanation works, it would not matter how causally heterogeneous (from the physicist’s point of view) the bases were; all that matters is that they produce the required effect. In some citations of functional properties it is obvious that the bases of the same functional property will be very different; one can say that the colour of the lion has the same function as the colour of the polar bear, which is also the same function as that possessed by the changing colour of the chameleon. This base variation will be important later. At the moment functional explanation is germane because it solves the harmony-of-effects problem: the instantiation of the functional property ensures the production of the effect because it is instantiated in a base selected precisely because past instances of that base produced that type of effect.

13 The example is drawn from The Problems of Evolution by Mark Ridley, Oxford University Press, 1985, p.69.
III

It is important to note that natural selection in biological contexts is only one way of solving the dual problem of why the supervening and base properties coincide, and why they produce the right effects. In more general terms, any supervening property which is 'effect-sensitive', i.e. relationally specified with respect to its effects, will be appropriately coincident with the base properties.Dispositionally defined properties are a prime example, and it may be thought that the scientific naturalist could rest content with a solution to the problem of supervenient causation which simply treated all causally efficacious supervenient properties as dispositional, and the relevant explanations as dispositional explanations.14 The relevant problem here is that this is at least insufficient in this context, for three reasons. First, the domains in which we are interested are those in which there is a set of properties which seem to be unified and set apart from every other domain of properties: biological properties appear to be distinct from physical-chemical properties; these also seem distinct from psychological properties, which in turn are distinct from social properties. Treating all these properties as merely dispositional does not solve the problem of what differentiates them. It may be argued that nothing does divide them, but that in effect denies the common assumption that there is a problem connected with the causal power of supervenient properties.15

The distinctiveness of the causally efficacious properties carries over to distinctiveness of the effects produced. On a 'mere' dispositional account, the disposition is construed as a higher order property which, when instantiated, produces the lower order effect. In the domains of interest to the scientific naturalist this is typically not true: the higher order property produces a higher order effect,


15 Tim Crane and D. H. Mellor argue that modern physics falsifies the supervenience picture of the social sciences, and so reject the idea that there is a problem about their causal power. See their 'There is No Question of Physicalism', Mind, Vol. 99, No. 394, April 1990, pp. 185–206.
except in cases of causal interaction, which are not central. This is best seen in the psychological case, where the intentional properties produce actions, not just bodily movements. In the biological case it is more difficult to discern this difference, but the (ultimate) end of functional properties is to produce relative reproductive advantage, a biological property.

This leads on to the second, related, reason for the insufficiency of the dispositional solution, which is that denying that there are relevant distinctions between the different domains is suggestive of reductionist solutions, and usually what blocks proposed reductions is that the specific set of normative commitments which are associated with the relevant domain cannot be captured at the reducing level. The biological properties are functional properties, and this provides the setting for judgments about malfunctions, or a dysfunctioning system of properties. Similarly, psychological properties (also relationally defined in an effect-sensitive way) are holistically united by norms of rationality, which cannot be captured in a biological-functional system nor in a nomological causal system. Straightforward dispositional accounts have great difficulty in capturing these norms. (Even sophisticated dispositional analyses fail this test: the 'directively organised systems' approach to the functionality of biological properties makes it very difficult to see how the notion of malfunction can get a grip. 16)

These normative commitments are important for the question at issue, the causal potency of the supervening properties, because they contribute to the determination of the causal shape of the supervening property. The effects which the supervening property is disposed to bring about are those which their normative character implies they are supposed to bring about. So the causal power of a biologically functional property may diverge from that of the base

16 See Alexander Rosenberg's account of 'function' in his The Structure of Biological Science, Cambridge University Press, Cambridge. For an elaboration of these remarks, see Graham Macdonald, 'Reduction and Evolutionary Biology'. Functionalism in the philosophy of mind is also a sophisticated dispositional account of psychological properties, and is criticised for the same reason. See John McDowell, 'Functionalism and Anomalous Monism' in E.Lepore and B.McLaughlin, eds., Actions and Events, Blackwell, Oxford, 1985, pp.387–398.
which instantiates it on any one occasion; it will be tied to the effects it is meant to produce, *qua* functional property.

The third reason for avoiding the dispositional accounts is that they positively cry out for elaboration at the causal level: they appear to be incomplete unless causally filled out. This invites the understanding that the supervenient explanations are stop-gap, second best explanations which will be supplanted by the causal explanations when they are available. If this were not so we would have two independently viable, non-normative causal explanations of the same phenomenon, a puzzling state of affairs at least. This is exacerbated by the thought that dispositional properties must be *locally* causally supervenient; insofar as our supervenient properties are not, then the simple dispositional account of their causal power will be revisionary of our normal understanding of those properties. The non-revisionary Wittgensteinian will do well to be sceptical of such dispositional accounts.

The pure dispositional approach, insofar as it invites this causal completion, separates the issues of normativity and explanation; it treats the normative characteristics of a supervenient system as though they were tacked on to the dispositional explanations as an optional extra. On the approach which is being recommended here, the normative consideration will arise directly as a consequence of how the effect-sensitive properties come to have their effect sensitivity. The selectionist story shows how certain causal histories produce functional, hence effect-sensitive, properties. On this view, the hope for a naturalistic psychology is *not* that intentional properties will figure in causal laws, but that just as the functional emerged from the causal, so intentional properties will be seen to have emerged from functionality.

It is worth remarking that it would be counter-intuitive to insist that the instantiations of dispositional properties, or of the programme properties and biological properties, were not causally efficacious just because they were instantiations of higher order properties. But that is a position which some have adopted; Jackson and Pettit view such properties as non-causally bringing about the

---

17 It is so puzzling to some that they wish to eradicate this possibility altogether. See Jaegwon Kim, 'Explanatory Exclusion And The Problem Of Mental Causation' in E.Villanueva, ed., *Information, Semantics & Epistemology*. 
instantiation of the causally efficacious properties, which then causally produce the required effect. At other times they suggest that even if one did take the view that the instance of the supervening property was causally efficacious, it would be so only in a derivative way. This indicates that their denial of genuine causal efficacy is related to our second worry: genuine causal power is not properly located at the level of the instance of the property. Take their definition of a causally efficacious property:

A causally efficacious property with regard to an effect is a property in virtue of whose instantiation, at least in part, the effect occurs; the instance of the property helps to produce the effect and does so because it is an instance of that property.

The intention is to show that a higher order property, say the property [F], is not causally efficacious because even if an instance of it, [iF], ‘helps to produce’ the effect instance, [iE], it does not do so ‘because’ it is an instance of [F]. So [F] is not causally efficacious, but is causally ‘relevant’. Note that here Jackson and Pettit are not denying that [iF] is causally efficacious: in the case where [iF] = [iG], they say that ‘if both are efficacious, then [F] [the higher order property] is efficacious only in a derivative sense. The relation between the instantiation of F and the occurrence of e is secondary to the relation between the instantiation of G and that occurrence; other things being equal, the obtaining of the latter relation ensures the obtaining of the former, and not vice versa.’

However, where [iF] = [iG], the obtaining of [F], (i.e.,[iF]) ensures the obtaining of [G], (i.e.,[iG]) just as surely as vice versa, and so is equally efficacious in ‘ensuring’ the relation to the obtaining of e. Jackson and Pettit go on to say that any efficacy which [F] is alleged to exercise would be ‘screened off’ by the influence of [G]. Again

18 As well as ‘Program explanation: a general perspective’, see ‘Functionalism and Broad Content’ *Mind*, vol XCVII no. 387, July 1988.
20 *ibid*, emphasis added. The ‘and not vice versa’ is in flat contradiction to what they say elsewhere, where it is claimed that the instantiation of the higher order property ‘ensures’ the instantiation of the lower order property. See *ibid*, pp.114–5, and in ‘Functionalism and Broad Content’, p.395. Their examples tend to tell against the ‘higher to lower’ direction of the ensuring relation. Take ‘Someone’s coughing’ and ‘Fred’s coughing’. The first is entailed by the second, not vice versa.
this is not possible where the instances of the properties are identical.

The best route for the causal impotence theorist to take would be to deny the coincidence of the properties. But if the instantiations of the higher order property were separate from those of the lower order property, then we would be owed an explanation of how they are related: how it is that the one non-causally brings about the other. We are not denying that these properties are non-causally related, but the best way to make sense of this is to provide an account of how it is that one instantiation non-causally ensures the instantiation of the second. Again, the question one has to ask is: how is it that co-occurrence of the relevant properties occurs? The most plausible answers would appear to be those which ensure co-instantiation. Obvious candidate answers are that the properties are logically or metaphysically related, or that some kind of empirical explanation is available.21 The first two provide grounds for the coincidence view, so an empirical account would be required. Reduction would be one such explanation, but that both ensures coincidence and goes against the supervenience assumption. The one we have provided, selectionism, also leads naturally to coincidence and allows for supervenience. Without an alternative account of co-occurrence, coincidence seems assured.

We have seen that the causal impotence view allows for coincidence but still denies direct efficacy. The underlying problem here is that there is a preference for defining causal efficacy with respect to properties. This involves large issues which cannot be settled here, except to say that given it is instances which produce effects, the definition of ‘causally efficacious’ should apply in the first place to instances. If one starts with properties, then one way forward may be to claim that a property is causally efficacious when it is used essentially in a causal explanation of an event, in particular in that part of the explanation which makes essential use of causal laws.22 This route confuses the (causal) explanatory issue with the

21 The empirical explanation must be suitably ontological: relying on epistemological answers, such as inductive relations, would be to rely on the wrong kind of explanation. Similarly, saying that the property instances are connected by law would be to cite the wrong kind of connection, at least on a realist reading of laws whereby they are true in virtue of worldly matters (properties and relations).

22 This is not available to Jackson and Pettit, though, on the grounds that they insist that the
ontological issue of causal efficacy, and makes causal efficacy an intensional rather than extensional relation. This implausible move is unnecessary, given the alternative sketched above. The conclusion is that, short of ruling out the possibility of distinct properties being co-instantiated, if we respect the distinction between causal explanation and causal efficacy, then the supervening properties are causally relevant because they are causally efficacious.

(It is certainly true that not all supervening properties are causally efficacious just in virtue of being instanced: the heat of the coal, not its redness, causes the burn. So work does need to be done on which property, supervening or otherwise, is causally efficacious, but this is likely to be empirical work. Remember here the nature of our problem, that it is one to do with the possibility of a supervening property being causally efficacious. The two questions, how supervening causal efficacy is possible and which properties are causally efficacious, must be kept apart. The co-instantiation solution is an answer to the first, not the second, question.)

IV

The scientific naturalist’s concern with locating supervenient properties in the causal network seems to lead to the conclusion that supervenient explanations must be restricted to causal explanations. What has been suggested above is that the best way of understanding supervenient and subvenient causal harmony is via understanding how the supervening properties come to be relationally defined (more specifically, relationally defined with respect to their effects). This in turn gives rise to a type of explanation which exhibits characteristics different from those of nomological causal explanation: accompanying types of effect-sensitivity is a normativity peculiar to that type, and the causal bases of the supervening causal type can vary. Now it may

higher order property does figure in causal explanations of events; and there would be a manifest circularity if they alluded to the different ways (causally relevant vs. causally efficacious) in which the property was used in causal explanations.

23 The example is due to Gabriel Segal and Elliot Sober, in ‘The Causal Efficacy of Content’, Philosophical Studies, 63, pp.1–30.
well be that the typical scientific naturalist is happy to accommodate these differences and thinks that whether one calls such cause-citing supervenient explanations nomological-causal or not is simply a terminological quibble. This is fine, except that some hefty conclusions for folk psychology seem to accompany the commitment to causal explanation, conclusions which gain their plausibility from reading that commitment in the narrow rather than the extended sense. In particular, arguments which lead (a) to eliminativist conclusions, (b) to the conclusion that there must be a Language of Thought (hereafter LOT), and (c) to the claim that a scientific psychology must taxonomise in accordance with narrow rather than broad content, presuppose substantial constraints on causal explanation. I will look at these in turn.

(a) One way in which folk psychology is challenged is *via* the studies which attack the assumed modularity in the organisation of our beliefs. Stephen Stich claims a belief store is modular if 'there is some more or less isolatable part of the system which plays (or would play) the central role in a typical causal history leading to the utterance of a sentence'. If there is no such isolatable belief state which plays the right causal role, then it appears that a central assumption about belief states will be wrong. As a consequence, so Stich says, belief ascriptions will typically lack a truth value. It turns out that there are models in which no single causal component ('or naturally isolatable part') underlies the expression of belief. If these models turn out to be the best available, so much the worse for belief states.

In similar vein it is argued that a certain view of connectionist models of the mind would have the same consequences for belief. There is a distinct disanalogy, it is said, between these distributed network models of, say, memory, and some semantic network models. In the latter, for any proposition stored in memory one can locate a distinct functionally discrete sub-part of the system associated with that proposition. Thus when folk psychology judges

---


that a certain belief has been activated in the process of inferring to a conclusion, there will be a correct answer to the question of whether that judgment is correct. It will be correct iff the belief-associated sub-part was causally active in that process. By contrast, given that there is no such discrete sub-part underlying any of the propositions which the connectionist system can be said to be storing, the question of which belief was causally active in the production of a conclusion cannot be answered. Here again the modularity assumption, said to be a prominent part of folk psychology's conception of what it is to have a belief, is false.

What is doing the work here is the assumption that folk psychology is committed to the thesis that for any belief there is the isolatable causal sub-part of the system with which that belief is associated. It is difficult to know how to verify the claim that this is true of the folk psychological conception of belief; for what it is worth, most students do not seem resistant to the idea that beliefs can have their effects via causal routes which can be various. More importantly, though, the considerations adduced above suggest that it is an unreasonable demand to place upon a supervenient system that it has parts which are so tightly connected to the causal sub-structure. The demand is that whenever a belief is activated in the causation of an action there is a central role played by the associated part of the sub-structure in the causal history of that action. The same demand in biology would be that whenever an effect occurs as a consequence of the exercise of a function there is one isolatable part of the system which is causally responsible. The example of protein synthesis used above is a warning that this need not be the case. In general, nature tends to be causally profligate with functions: if a function is important it is likely that there will be back-up systems which will perform the same function in different ways in order to safeguard against the failure of one of the mechanisms responsible for one of the ways of exercising that function.26

(b) Similar causal requirements play a part in the arguments supporting the 'language of thought' hypothesis. Take one of Fodor's arguments, the Methodological Argument.27 It relies on

26 Ruth Garrett-Millikan makes this point in her 'Explanation in Biopsychology'.
principle P which claims that if two causes $c_1$ and $c_2$ normally have effects $e_1$ and $e_2$ respectively, and if $c_3$ normally has effects $e_1$ and $e_2$, then it is reasonable to infer that $c_3$ has as its constituents $c_1$ and $c_2$. Applied to mental states the conclusion is that such states have constituent structure in virtue of the complex events they cause; and such structure is what the LOT hypothesis is claiming.

The justification for principle P is interesting. The suggestion is that if we reject it we are committed to the ontologically excessive claim that the same event, say $e_1$, can be brought about in distinct ways, one way when it is caused by $c_1$ and a different way when it is caused by $c_3$. But this is precisely the possibility recognised by the idea that (causally efficacious) supervening properties are variably realisable; it is also the possibility which is guarded against by the claim that the supervening properties are not picked out in virtue of their viability in causal-nomological laws. Were these properties to be candidates for use (or mention, depending on your account of laws) in such laws, then the reductionist import of principle P, and the LOT hypothesis, would be justified: variable realisation would be a mystery.\(^{28}\) Given that the supervening taxonomy in this case is determined by intentional (broadly construed) considerations, there is simply no presumption that the intentional or semantic structure should be copied onto the causal base.

(c) Consider the following argument, again due to Fodor:

1. My twin and I are molecular duplicates.

2. Therefore our actual and counterfactual behaviours are identical in relevant respects.

3. Therefore the causal powers of our mental states are identical in relevant respects.

4. Therefore my twin and I belong to the same natural kind for purposes of psychological explanation and ‘individualism’ is true.\(^{29}\)


28 This is argued in the early sections of ‘The Possibility of the Disunity of Science’. There is simply a tension between Fodor’s antireductionism (exemplified in ‘The Special Sciences’ *Synthese*, 1974) and the reductionism of the LOT hypothesis.

Here once again the example of biology should prevent us from accepting the conclusion. In essence the argument stipulates that all natural kinds (for the purposes of psychological explanation) must supervene on local (physical) causality. Biological (functional) explanation uses ‘kinds’ which do not so supervene: molecular duplicates may well be functionally different, given different histories. Say one twin has its properties as a consequence of an evolutionary history featuring natural selection for (some of) those properties, whilst the other twin is an instantaneous creation. From the perspective of physics the causal powers of the molecular twins must be the same. From the perspective of biology, the behaviours of evolved twin may require a different explanation from those of instant twin. The first may be performing biological functions, the second will not. Given the importance of selection processes to biological taxonomies, it would be difficult to claim that the different supervenient explanations associated with the molecular duplicates operated only at the level of ‘folk’ biology. The moral is that psychological patterns will probably carve the world into different kinds from those determined by local causal supervenience. To deny that these are natural kinds is simply a prejudice in favour of one particular pattern, the causal-nomological.

V

What has been argued for above is that the problem of supervenient causation can (and should) be settled in a way which leads to the conclusion that the types of explanations appropriate to a supervening science will be those which eschew the use of causal-nomological laws. One can have supervening causally efficacious properties whilst recognising that there are more patterns to be found in nature than just those of the regular causal kind. As a consequence some of the problems which have been thought to plague attempts to naturalise the mental, to take one

30 This formulation leaves it open as to whether the twins must share all their causal powers. If one is committed to the view that they must be causally indistinguishable, then the claim here is that causal indistinguishability does not prohibit supervenient property differences, where those property differences are ‘scientifically’ relevant.
example, are seen to arise out of a prejudice which restricts the proper form of scientific explanation to causal explanation, where this has a more substantial import than simply the citation of causally effective properties.

The connection with the Wittgensteinian form of naturalism lies in the non-revisionary aspect which emerges from the considerations about causality. Revisionism, reductionism, and eliminativism in philosophy of mind have been, at least in part, motivated by attempts to align mental causation with physical causation or, if this is thought impossible, to revise our conception of the psychological in the light of this failure. The analogy with biology is useful insofar as it provides an example where the causal considerations lack any bite, and do so for the abstract reasons set out in the consideration of supervenient causal properties: what is required is a form of property which is specified relationally. The intrinsic states require extrinsic relata. The position which emerges is consonant with that expressed by Wittgenstein in his rejection of a Language of Thought:

No supposition seems to me more natural than that there is no process in the brain correlated with associating or with thinking; so that it would be impossible to read off thought processes from brain-processes. I mean this: if I talk or write there is, I assume, a system of impulses going out from my brain and correlated with my spoken or written thoughts. But why should the system continue further in the direction of the centre? Why should this order not proceed, so to speak, out of chaos?31

In his commentary on this passage Malcolm Budd says:

The moral is clear. Wittgenstein is claiming that it is conceivable that the causal chains inside a person’s body that bring about the bodily movements in which his thoughts are expressed might terminate in chaos.

and later:

Wittgenstein wishes to allow that different counter-factual conditionals could be true of people in the same circumstances,

without the difference being dependent upon a physical difference in the conditions of the people's bodies.  

This is what has been argued for above. This is not to say that all the problems associated with naturalising the mental, say, have been solved. A more modest conclusion is recommended: one can have causal efficacy without mirroring physical causality, with physical 'causal chaos'. A certain kind of revisionism can be avoided if one pays attention to the different kinds of explanatory language games that there are in science. Wittgenstein's aversion to 'theory' can then be interpreted as an aversion to the adherence to one single type of theory, that in which one uncovers the hidden (causal) essence of the process under study. Different types of process can require different types of explanation.

Paradoxically, recognising this at the same time renders it less important whether one calls the explanations 'scientific' or not. One can choose to apply the term to the single pattern, leaving biology, psychology, and sociology beyond the pale; but there will be no motivation stemming from concerns about causality for this decision. There will be even less motivation for insisting that what is beyond the pale is unnatural. Any major remaining difficulty of naturalising psychology will have to do with (at least) the exact nature of the relation between the biological and the mental, where this will incorporate problems associated with sub-doxastic states and the differences, or similarities, between rational and functional patterns.  

None of this, it seems to me, will threaten this tentative move to a reconciliation between Wittgensteinian and scientific naturalism.

---


33 My own prejudice is in favour of emergent mentality—a non-reductionist position.
THE NATURE OF NATURALISM

Graham Macdonald and Philip Pettit

II—Philip Pettit

I

Introduction. Naturalism imposes a constraint on what there can be, stipulating that there are no nonnatural or unnatural, praeternatural or supernatural, entities. It raises a challenge, therefore, for the beliefs we spontaneously hold, inviting us to ask whether they are compatible with this constraint and, if not, whether they can be rendered compatible with it. The naturalistic program in philosophy involves the attempt to meet that challenge.

The naturalistic program drives much contemporary research in philosophy. Consider the debates about secondary qualities and values, about natural laws and modalities, about qualia and free will and folk psychology, about personal identity and the flow of time. The main issue in all of these discussions is whether nature allows us to maintain true beliefs involving such matters. Do the beliefs incur ontological debts which nature will not honour? Do they require ontological resources beyond anything that nature can provide? Or can we retain the beliefs, perhaps after a little reworking, within an explicitly naturalistic economy?

The main divide in such discussions is between soft naturalists and hard naturalists. Hard naturalists deny that the naturalistic constraint can be satisfied in this or that area of belief and so they choose to reject the spontaneous appearances as mere illusion. They argue that ordinary objects are not really coloured, or they repudiate our folk psychology of mind, or they hold that value-discourse is purely expressive of feeling: the subjects debunked, and the debunking strategies, are legion. Soft naturalists are more accommodating. They argue, to the contrary, that the naturalistic constraint is satisfied in this or that contentious area and that the appearances can be saved there after all. They make the case for a
naturalistic realism about colours or mental states or value-properties or whatever.

Our concern here is with the nature of naturalism. In order to elucidate the topic, three things are necessary. First, to appreciate the sort of constraint which naturalism imposes; second, to identify the sorts of challenges which it raises; and third, to see the sort of thing that is required if such challenges are to be met. I discuss the naturalistic constraint in the second section; I distinguish two sorts of challenges in the third section: I call these the 'eclipsing' and 'fading' problems; and then in the remaining parts of the paper I discuss the first and less familiar of these challenges in some detail. This challenge is the main focus in Graham Macdonald’s paper and I discuss that paper in dealing with the challenge.1

II

The naturalistic constraint. Extra naturam nulla salus. Outside nature no salvation. If we want a slogan, then this will do as well as any to define the naturalistic constraint. It dictates that among the appearances countenanced in any human experience or activity, only those that pass a certain test can be saved: only those that prove to reflect some aspect of the natural world. The principle challenges us to examine a variety of otherwise persuasive appearances to see how they can pass this test. It challenges us to look at whether, as naturalists, we can save the appearances associated, for example, with talk of the mental, the modal and the moral; talk of objective colour and personal identity; talk of the passage of time and the solidity of middle-sized objects; and so on across a familiar spectrum of topics.

The naturalistic slogan is subject to different interpretations. It leaves it unclear as to what is comprised by natura or nature; it leaves it undetermined as to what is to count as extra or outside; and it does not say what constitutes salus or salvation. I cannot discuss these indeterminacies fully but I must at least give a rough indication of how I think that they are best resolved.

1 'The Nature of Naturalism', this volume. The differences between us, or at least the differences that can be usefully debated in a short discussion, all turn on this matter.
Nature comprises those entities, and constructs made of those entities, that the ideal physics, realistically interpreted, posits. As envisaged here, physics aspires to explore the microconstituents out of which everything in the spatio-temporal world is composed and to identify reliable regularities sufficient to account for the behaviour of those constituents. It aspires to be comprehensive in scope and complete in coverage. Unlike chemistry and biology, for example, physics does not confine its attention to things that are large enough to have a chemical structure and organised enough to be living. Thus it has a special place in the identification of what belongs to nature. Or so at least the naturalist holds.

What are the sorts of entities likely to be posited in the ideal physics, realistically interpreted? There are a variety of views. Inflationary accounts would include universals, causal or nomic connections, objective chances and even possible worlds, among the primitive entities recognised. More deflationary reckonings would reduce the primitive posits to particular space-time points, particular bits of matter, and the intrinsic, physical properties of such entities: properties like spin or charge, or properties introduced by the ideal physics to derive or displace such currently familiar properties.2

Now to the second indeterminacy in our slogan. What is it for an alleged entity to fall outside nature? What is it for an appearance not to reflect an aspect of the natural world? There are two possibilities to be distinguished. One is where the existence of the apparent entity would require something beyond the sort of stuff posited in physics: something beyond space-time points and bits of matter, for example, if those are taken to be the only primitive entities that physics recognises. The other is where the existence of the apparent entity does not require non-physical constituents—it is not a Cartesian mind, for example—but where it can apparently cease to exist, or change its character, without any change of a kind that physics can register. The apparent entity is composed of physical constituents, as we might put it, but it is not subject to any constraint of physical supervenience. The fate or character of the entity may vary while things remain unchanged in the perspective

of physics: in the perspective where we register everything relevant to the behaviour of the thing’s ultimate constituents.3

What is it, finally, for an appearance or set of appearances to be saved? What constitutes the salvation mentioned in our slogan? If we think of the appearances in any area as corresponding to a discourse, then we may say that they will be saved just in case the discourse involved lends itself to a realistic construal. The discourse directs us to entities that are distinctive: they are of kinds which are not \textit{a priori} identifiable, one by one, with kinds of thing recognised elsewhere; those entities exist and exist independently of our recognition of them; and they are entities about which we are liable to fall into error and ignorance.4

I cannot elaborate on this account of the naturalistic constraint but I should remark on one feature. The constraint is made less demanding than it might be by the manner in which the second indeterminacy is resolved. Anything that is physical in composition and that satisfies physical supervenience will count as naturalistically acceptable under the preferred way of resolving that indeterminacy. Thus the approach leaves room for entities that are quite alien to the posits of physics and the other sciences; it leaves room for entities which are not invoked, for example, to any scientific purpose. Naturalism is probably better described as ‘scientism’, if it is taken to require that any entities countenanced are essential to some scientific theory. Our characterisation of naturalism is decidedly non-scientistic.

III

\textit{Two naturalistic challenges.} Under our account, and under any plausible representation, the naturalist endorses a more or less austere account of the natural world: roughly, an account within the atoms-in-the-void tradition of thinking. This means that he is

3 The supervenience requirement does not make the composition requirement redundant. Numbers are bound to satisfy the supervenience requirement, for example, since mathematical truths are supervenient on all other truths—this is because they hold necessarily—and a platonist view of numbers offends only against the composition requirement. On these matters see G. Hellman and F. Thompson ‘Physicalism: Ontology, Determination and Reduction’, \textit{Journal of Philosophy}, Vol. 72, 1975.

challenged to square that image of the natural world with the image of the world presented via less theoretical routes. In the terminology of Wilfred Sellars, the challenge is to try and reconcile the scientific image of the world with the manifest image.\(^{5}\)

There are two distinct sorts of problems which the naturalist is likely to encounter if he takes up this challenge and pursues a mission of reconciliation. The first is an eclipsing problem, as I will call it, the second a problem of fading.

In the scientific image of the world a variety of properties will be ascribed to whatever entities are countenanced. For example, if the scientific image recognises certain microphysical particles, it will ascribe to them properties like mass and charge and spin. With certain properties, however, the fact that they are ascribed to the basic entities involved in the scientific image may seem to undermine the possibility that they equally belong, as they seem to belong, to the less basic entities countenanced in the manifest image. The fact that they belong to the microphysical entities out of which everything else is composed may eclipse the appearance of their belonging to anything less fundamental. It may make it difficult to preserve that appearance intact.

This eclipsing problem arises with properties related to cause and effect, and lawlike connection: for short, let us say, with dynamical properties. Suppose that our picture of nature, our scientific image of the world, is one which allows that microphysical entities, say microphysical particles, are causally relevant—it does not matter in what sense—to what happens from moment to moment. Let us say that the picture identifies certain properties of those entities as playing a directive—though perhaps just probabilistic—part in the evolving history of the world. The admission of such causally relevant properties may have the effect of compromising the ascription of causal relevance to any other properties, in particular to any properties identified only in the manifest image. The presence of causal relevance in the scientific image, as we can put it, may eclipse the appearance of causal relevance in the manifest.

The eclipsing effect is supported by a beguiling line of argument. Everything that exists is composed of the stuff that physics

identifies and whatever happens to anything is supervenient on changes that physics can track. Or so at least the naturalist holds. But this means that if certain physical properties play a directive role in the evolving history of the physical world, then they play that role relative to everything that occurs. Nothing happens without a physical cause. Nothing happens without an ancestry in which certain physical properties figure as causally relevant.

Consider something like a psychological change or a change of colour or even a change in the value of an object. That change is supervenient on microphysical changes and certain microphysical properties are causally relevant in the generation of those changes. But this means that the microphysical properties are also causally relevant in the generation of the non-microphysical change. For causal relevance, intuitively, is transitive across the supervenience relationship. If the hemlock is causally relevant to the death of Socrates, then it is causally relevant to the widowing of Xanthippe. If the humidity is causally relevant to the change of colour in the painting, it is causally relevant to the associated aesthetic loss.

Now to the eclipsing effect. If everything that happens has a microphysical cause, then a question arises as to whether there are any other causes. If microphysical properties are causally relevant in the determination of microphysical happenings, and relevant thereby in the determination of everything that happens, then where is the room for causally relevant non-microphysical properties? We might say that such properties overdetermine the effects of the microphysical properties but that hypothesis looks entirely ad hoc. The convinced naturalist is going to be tempted by the thought that only microphysical properties, properties of the kind that show up in physics, are causally relevant and that the appearance of causal relevance among other properties is an illusion. He must be inclined to let physical causal relevance eclipse the causal relevance that is sometimes manifested by non-microphysical properties.

The cost of allowing this eclipsing effect will be considerable. We have no hesitation about ascribing causal relevance to chemical

and biological and psychological properties, for example, and if we allow the eclipse of non-microphysical causality then we must inhibit our inclination to see causal relevance in these realms. We must cease to think that it is in virtue of our believing and desiring various things that we do what we do, for example. We must recognise that the cause of our action lies hidden in the microphysical infrastructure of our psychology and that it is never going to be revealed in the perspective of psychological concepts. Whatever patterns those concepts allow us to discern, the entities to which they direct us are entirely impotent and inert. The psychological states that we pick out in the lead-up to action are no more generative of that action than a film clip is generative of the clip that follows.

So much for the eclipsing problem that confronts naturalists. The fading problem is a complementary sort of difficulty. Where the eclipsing problem arises because of what is present in the physical realm—causal relevance or whatever—the fading problem arises because of what is absent in that realm. Where the eclipsing problem is generated by the richness of physical nature, the fading problem is generated by its poverty.

The fading problem represents a familiar challenge. If the nature recognised in physics is colourless and odourless and tasteless, then how can there be secondary qualities of the kind to which our senses attest? If that nature is valueless, how can there be room for the values from which we take our guidance, or at least our counsel? If that nature is lacking in consciousness and intentionality, then how can we make place for the manifestations of mentality which we find in ourselves and others? Such appearances, though they are the very stuff of our ordinary experience, seem to fade away in light of the awful austerity of atoms-in-the-void. The entities to which they testify begin to look like will o’ the wisps.

Under our account of the naturalistic constraint, the essential trick that needs to be turned with such problematic entities is to show that they are microphysically constituted and microphysically supervenient. We will have to be able to argue that they do not require any stuff beyond the primitive posits of physics, be those space-time points, bits of matter, or whatever. And we will have to be able to make a case for thinking that the fate and character of the entities is superveniently fixed by what happens at the
microphysical level. Otherwise the fading effect may banish the entities from our picture of what there is. The fading problem should be familiar, for much of contemporary philosophy is concerned with arguing for the required sort of composition and supervenience in this or that area. There are a number of strategies for dealing with the problem but this is not the place to review them. It would require a treatise to do them justice.  

This should be enough by way of presenting the two sorts of problems which the naturalist confronts. I now turn to how the naturalist might think of dealing with the eclipsing problem. Where the fading problem is at the focus of much contemporary research, the eclipsing problem has attracted surprisingly little attention.

IV  
The eclipsing problem. Suppose that certain microphysical properties are causally relevant to a given event, an event which is constituted out of microphysical changes and which is supervenient on those changes. The event may be a change of colour or shape, it may be an action or a change of mind, it may be the movement of a ball or the death of a person. With any such event we identify certain non-microphysical properties as causally relevant: the heat we applied changed the colour or shape of the object, the pressure caused the ball to move, the knife in the back caused the person’s death. The eclipsing problem has to do with how we can ascribe causal relevance to such properties—if you like, to the instances of those properties—consistently with taking the microphysical properties also to be causally relevant.

The microphysical properties relate to the properties invoked in these examples as properties at a lower level relate to properties at a higher. Levels are characterised by families of properties such that properties in two families can be causally relevant to the same event without combining as sequential or co-ordinate causal factors: without the property-instance from either family being a cause of

7 For the record, my own preference is for the so-called ‘ethocentric’ strategy described in ‘Realism and Response-dependence’. My sense of the possibilities has been shaped by collaborative work, with Peter Menzies, on how far that strategy can be exploited in the service of a soft naturalism.
the presence of the other; and without the property-instances combining to produce the event in the way the striking of the match and the presence of oxygen combine to produce flame. When is one level, one family of properties, lower or higher than another? Assume that properties from the two levels are relevant to the same event, as in the possibility just envisaged. Intuitively, the property from the lower level will mediate the causal relevance of the property from the higher, and not vice versa.8

Imagine that the malleability of this eraser and its exact molecular structure both help to produce a bending effect, as in the ordinary case: there are no deviant causal chains at work, for example. The properties are at different levels because, first, neither is causally relevant to the presence of the other; and, second, the properties do not causally combine to produce the bending, as if they represented different causal requirements. Which level is higher, then, which lower? The malleability property is at a higher level than the molecular-structural properties because, intuitively, its causal relevance to the bending is mediated by the molecular-structural properties, and not vice versa: there is no way that the malleability can be relevant to the bending of the eraser except through the molecular-structural properties being relevant, and not vice versa. Or so at least most of us will happily assume.

The core of the eclipsing problem is to explain how properties at lower and higher levels can collaborate in the production of an event and can collaborate, in particular, so that both sorts of property count as causally relevant to the event. Specifically, it is the problem of explaining this multiple relevance in the typical case, when the

---

8 This account of levels rests on intuitions about the absence of sequential or co-ordinate causal collaboration and on intuitions about the causal mediation exercised by certain properties relative to others; furthermore, the account is theoretically controversial, for the intuition about mediation may be resisted by the outright singularists mentioned later. But none of this need concern us. The naturalistic eclipsing problem—and more generally, as we shall see, the problem which the program model tries to resolve—is precisely the problem of saving intuitions such as those which are built into this account of levels: intuitions some of which may be rejected by various theorists. Presenting the problem in terms of levels is merely an attractive expository strategy. It offers a way of coding the five intuitions resolved under the program model: that a certain property-set is causally relevant to an event; that a second property-set is relevant also; that the properties do not combine sequentially to help produce the event; that neither do they combine collaterally to do so; and that one of the properties mediates the causal relevance of the other.
properties at the different levels are not themselves causally connected as parts of the same causal complex or chain. If we can solve this problem, then we will have no difficulty in seeing how properties recognised in the manifest image of the world can be causally relevant to an event, even though the event is the upshot of a microphysical causal history. We will have no difficulty in resisting the eclipsing effect and in saving the appearance of causal relevance outside the microphysical realm.

The program model that Frank Jackson and I have developed purports to resolve the problem of how properties or property-instances at lower and higher levels can be causally relevant to one and the same singular thing. It abstracts from how causal relevance is to be understood—causal relevance is taken to be a matter of intuitive judgment—and focuses on the way relevance, however paraphrased, may be reproduced across levels. Suppose that there is no doubt about the causal relevance of properties at a given level $L$ to the occurrence of an event $E$, of a given type. Suppose that we are interested in how a property, $P$, at a higher level can be simultaneously relevant to $E$. According to the program model, $P$ will be causally relevant to $E$ just in case three conditions are fulfilled.

The conditions are these.

1. The instantiation of $P$ non-causally involves the instantiation of certain properties—perhaps these, perhaps those—at the lower level $L$: typically, the instantiation of the $L$-properties will 'realise' $P$, as it is said, given the context.

2. $L$-properties of the sort associated with instantiations of $P$, or at least most of them, are such as generally to be causally relevant—in the circumstances—to the occurrence of an $E$-type event.

3. The $L$-properties associated with the actual instantiation of $P$ are causally relevant to the occurrence of $E$.

---

These conditions are readily illustrated. Intuitively, the malleability of this eraser is causally relevant to its bending, and relevant simultaneously with the exact molecular structure. How so? Because the program model applies. The instantiation of the malleability involves the instantiation of certain molecular-structural properties; the sorts of properties associated with instantiations of malleability are such as generally to be causally relevant to the sort of bending effect in question; and the molecular-structural properties associated with the actual instantiation of malleability are causally relevant to the actual bending.\textsuperscript{10} I said that intuitively the malleability of the eraser is relevant to the bending, as are the molecular-structural properties, and that the program model makes sense of that dual relevance. But notice that even if the intuition about the malleability was not given, the program model would give us reason for assigning causal relevance to the property. The program model may help us in identifying certain properties as causally relevant—it may serve as a causal heuristic—even if its primary job is to take properties at different levels that are intuitively of causal relevance to something and to show how they can simultaneously enjoy such relevance.

Other examples of the program model become salient as we recognise suitably corresponding relations across levels in different cases. In every case the relation must be such that the instantiation of the higher-level property ensures or at least probabilifies—in a non-causal way—that there are causally relevant properties present at the lower level. But there may be quite different reasons applicable in the different cases as to why that relation obtains; each case will require its own annotation. The squareness of the peg probabilifies the sort of molecular contact which blocks the peg going through the round hole; the redness of the rag probabilifies the sort of physical stimulation which provokes the bull; the rise in unemployment probabilifies a shift in motives and opportunities

\textsuperscript{10} The example is more or less trivial, because knowledge that the molecular structure combined in the ordinary way with a certain outside force to cause the bending yields knowledge that the malleability was also causally relevant. But not all illustrations of the program model have this feature. This is noted at a number of places, for example in ‘Functionalism and Broad Content’, p.396. It explains why it is misleading to cast the approach as dispositionalist. Graham Macdonald comes close to suggesting such a representation of the approach in ‘The Nature of Naturalism’, fn. 14.
that is likely to increase aggregate crime; the rationality or perceived rationality of an action probabilifies the sort of neural disposition which produces suitable behaviour; and so on across a great variety of possible cases. The probabilification holds for different reasons in the different cases. But the fact that it obtains shows how the program model may apply in any of the examples, making sense of how the higher-level property can be causally relevant to something which is also traceable to the lower-level properties.

What alternatives are there to the program model? Consider again the case where properties at level L are agreed to be causally relevant to an event E. There are two salient, alternative accounts of what is necessary and sufficient for a property P, at another level, to be relevant to E. One would say that P is relevant just in case it is co-instantiated, as Cynthia and Graham Macdonald put it, with causally relevant properties at level L: just in case the instance of P is equally an instance of such properties. Another would say that P is causally relevant just in case it is supervenient on causally relevant properties at level L. The realisation of those properties ensures that P will be realised: realise those properties elsewhere in this world or in another possible world and, provided at least that certain other things remain the same, you will have realised the P-property as well.

Suppose that a piece of putty is resting on a metal mesh and, perhaps under the application of heat, it changes shape and falls

11 This model is defended, on at least one reading of the piece—a reading that takes causal efficacy to be the same as causal relevance—in Cynthia and Graham Macdonald 'Mental Causes and Explanation of Action', Philosophical Quarterly, Vol. 36, 1986, pp. 145–58. 'Just as to be red is to be coloured, one might say, to be an instance of the property, being a brain event B, can be an instance of the mental property, being a pain. Moreover, if an instance of the former is causally efficacious, then so is the latter'. See too Cynthia Macdonald Mind-Body Identity Theories, Routledge, London, 1989, pp. 163–64 and Graham Macdonald 'The Nature of Naturalism', where the same line is maintained. The model is broadly in the tradition of Donald Davidson's work but it figures also in many functionalist pictures of mind. It usually presupposes fulfilment of the supervenience constraint and so is more demanding than the second alternative.


13 There are details about the supervenience relationship that I ignore here, on the grounds that they are not relevant to my purposes. The proviso about other things remaining the same is not in Kim but it would appear that if his approach is to be plausible, then it has to be loosened on these lines; this is persuasively argued in Peter Menzies 'Against Causal Reductionism', Mind, Vol. 97, 1988, pp. 551–74.
through a hole in the mesh. The change of shape, we spontaneously say, causes it to fall through. Now we may assume in such a case that a change in the spatial arrangement of the microphysical parts of the putty is the corresponding cause at the physical level. So what is the relationship between the change of shape—the counterpart of the P-property—and the change in the arrangement of the physical parts—the L-change—which allows them both to be causally relevant to the putty falling?

The co-instantiation model suggests that the relationship must be one of identity: that there is just one event here, an event described at one level as a change in shape, at another as a change in the arrangement of the parts. The supervenience model holds that we do not have to posit an identity between the events: that we can think of them as distinct, given that the macro-event is suitably supervenient on the micro-event. That there is a certain change in the arrangement of the parts ensures in the supervenient way that there is a corresponding change of shape.

Neither of these models is satisfactory. The example given enables us to identify a problem which affects both. According to the first model of causal relevance across levels, the change of shape will be causally relevant to the putty’s fall just so far as that property-instance is identical with the change in the arrangement of physical parts that is relevant to the fall. According to the second model, it will be causally relevant just so far as it supervenes on the change in the arrangement of the parts. But suppose now that in changing the arrangement of its parts—say, under heating—the putty not only changes shape, it also changes volume: it expands slightly. If the first model identifies the change of shape with the change in arrangement of parts then it will also have to identify the change of volume with that change.14 And if the second holds that

---

14 This is a point at which the co-instantiation theorist can strain to resist the lesson of our example and of others like it. He may claim in every such case to have such a fine way of individuating the lower-level factors that he is not forced to hold that the two higher-level factors are each identical with it. Thus he may say that the change of shape is identical with the change in arrangement of parts, taken under one aspect—it had better not be the shape aspect, of course—and that the change of volume is identical with the change in arrangement of parts, taken under another. But this line will not be congenial to those who think in a relatively coarse way about property-instances, as co-instantiation theorists seem bound to do. Nor will it be congenial to those who are outright singularists, in the sense distinguished at the end of the paper, about the causal relevance of
the change of shape is supervenient on the change in arrangement of parts, then it will have to hold that equally the change of volume is supervenient. Thus each model will drive us to the conclusion that not only is the change of shape causally relevant to the putty’s falling through the mesh, so is the change of volume.

But that conclusion is outlandish. The change of volume has nothing to do with the putty’s falling through the mesh. The expansion of the putty makes it less likely, not more, that the putty will fall through. If it falls through as a result of changing shape, it falls through in spite of changing volume. Any model which forces us to regard the change of volume as causally relevant, and causally relevant in the same way as the change of shape, has to be inadequate.  

The program model, unlike the co-instantiation and supervenience models, has no difficulty in making sense of this example and of many others like it. The change of shape programs for the putty falling through the mesh, for it more or less ensures that the microphysical parts become distributed in such a way that there is no resistance to the putty’s falling through. The change of volume does no such thing. True, the way the putty actually changes volume is consistent with its falling through the mesh. But for the vast majority of ways in which that arrangement might change so as to yield the increase of volume, the putty will stay where it is. That the volume changes in the given way does not ensure or probabilify a change in the arrangement of the microphysical parts which is apt to let the putty fall.

In most of the cases where the program model applies it is plausible that there is a relation of supervenience, or at least partial supervenience, between the properties at level L which are assumed to be causally relevant to E and the P-property with which we are concerned. In that sense we may allow that the instantiation of those properties, given appropriate other conditions, ensures the property-instances; it makes instances all too sensitive in individuation to the properties of which they are instances.

15 The example used here to undermine the identity and supervenience models of causal relevance across levels is not particularly special. There are any number of examples like it and they all go to show that the models fail to give us a convincing picture of how non-physical factors come to be causally relevant to something. See Peter Menzies 'Against Causal Reductionism' for a variety of examples.
realisation of the P-property and not *vice versa*. But there is more than supervenience—or indeed co-instantiation—to the transmission of causal relevance. It must also be, as the program model has it, that the realisation of the P-property ensures or at least probabilifies the realisation of the L-properties: that is, it ensures or probabilifies the realisation of such L-properties as are apt to bring about E.\(^{16}\)

If the program model shows us how properties at different levels can be causally relevant to one and the same event or condition, then it shows us how to avoid the eclipsing problem. Naturalists grant that the microphysical properties discerned in physics are causally relevant to everything that happens. But it can still be the case that properties at other levels are causally relevant also. There is absolutely no need to be blinded to higher-level causal relations by the salience of lower-level connections.

V

*An objection.* This is not the place to argue again for the program model. But I should at least try to offer a defence against the epiphenomenalist charge which Graham Macdonald makes, when he describes the approach, by implication, as that of a ‘causal impotence theorist’.\(^{17}\) As I shall take the objection, the idea is that under the program model, the microphysical properties are causally relevant in a more fundamental way than any higher-level properties and that this asymmetry has the effect of undermining relevance at the higher levels. The model preserves the intuition that the microphysical properties represent a causally more basic level and in doing so it jeopardises the intuition that the higher-level properties have causal relevance. The approach means, in a phrase which Macdonald favours, that the higher-level properties are not, or at least need not be, causally efficacious.

\(^{16}\) No contradiction there, let alone the flat contradiction alleged by Graham Macdonald in ‘The Nature of Naturalism’, fn. 20. What Macdonald’s remark overlooks is the fact, central to the program model, that the P-property ensures the realisation of the L-properties in the sense of ensuring, not that precisely these L-properties are present, but that there are such L-properties present as are apt to bring about E.

\(^{17}\) ‘The Nature of Naturalism’, section 3.
The first point that I wish to make against Macdonald's criticism is that, for all that the program model requires, there need be no causally efficacious properties; among the properties that are causally relevant to a result, none need count as efficacious in any of the available senses. This point is relevant, because if the model does not strictly require any properties to be causally efficacious, then it can hardly be said to make microphysical properties efficacious in a distinctive and destructive sense.

There are three salient ways of understanding what it is for a property to be causally efficacious relative to a result E.

1. The substantive way. The property exercises a primitive, irreducible force in producing E.

2. The formal way. The property is causally relevant to E at the lowest level of causally relevant properties.\(^{18}\)

3. The easy way. The property is efficacious in sense 1 or 2 or is co-instantiated with a property that is efficacious in that sense.

Under the program model, there need be no substantively efficacious properties, as the model abstracts from whether such an irreducibilist theory of causation is sound.\(^{19}\) Equally, under the program model, there need be no formally efficacious properties, as it may just be that there is no bottom level: that, as some current physical theories maintain, there is an infinite progression downwards in the properties, and perhaps also in the particles, that might be regarded as causally relevant.\(^{20}\) Even if there is no bottom level, the program model makes sense of how causal relevance can be simultaneously found at the or at a microphysical level and at higher levels.

---

18 This, in effect, is how causal efficacy is understood in 'Program Explanation: A General Perspective'.

19 See 'Program Explanation: A General Perspective', pp. 108–9: 'the notion of efficacy it introduces is not tied to the view that causal efficacy is an irreducible feature of the world'.

20 See Ned Block 'Can the Mind Change the World?' in George Boolos, ed., Meaning and Method, Cambridge University Press, Cambridge, p. 168, who introduces this possibility in criticism, as he thinks, of the program model. The response made here was adumbrated in 'Program Explanation: A General Perspective', p. 116: 'if there is an infinite progression downward and therefore no efficacious properties...then the program story will have a different significance, bearing on relations between equally non-efficacious levels'.
The absence of a bottom level will have implications for how causal relevance at any level should be understood but, as already mentioned, the model does not presuppose any particular account of relevance. But notice that it would be mistaken to think that with an infinite progression downward, the model entails that there are no causally relevant properties at all. Consider an analogy. An aggregate-level object is positioned in space in virtue of the spatial positions of its parts as, under the program model, higher relations of causal relevance obtain in virtue of the obtaining of lower. But if there is an infinite progression downward in parts, that will not mean that there is no such thing as spatial position. And neither, by parallel, should such a progression downward be taken to mean that there is no such thing as causal relevance.

But suppose that there is a bottom level of causally relevant properties, as most naturalists probably believe. What then of the epiphenomenalist charge that under the program model, non-physical levels lose all causal potency? Does the charge gain any bite?

If there is a bottom level, then there are causally efficacious properties in the formal sense and the only such efficacious properties are the microphysical ones. But this need not be anything to strike a worrisome, epiphenomenalist note. Suppose that causal relevance is understood nomologically or in terms of necessary and sufficient conditions or in any fashion that does not paraphrase bottom-level causal relevance in terms of a substantive, irreducible efficacy. In that case higher-level properties will have causal relevance in just the same sense as microphysical properties and the only difference between the ways in which they rule the world, as it were, will be that the microphysical properties get in at a lower level: they are ontologically more basic, as the program model reveals. There is nothing to worry about here, and there is nothing to motivate talk of rendering the higher-level properties causally impotent. A little piece of fantasy may help to reinforce the point: in particular, it may help to show that the more basic status of

21 For example, in the manner set out in Peter Menzies and Huw Price 'Causation as a Secondary Quality', British Journal for the Philosophy of Science, forthcoming.
microphysical properties need not compromise the causal relevance of other properties.

Imagine a god-like designer who is concerned to establish macro-laws governing the aggregate shapes that succeed one another in a universe composed of little particles. If causal relevance is to be understood nomologically, or in a nomologically connected way, then we are to imagine this designer trying to set things up so that the appearance of this or that aggregate shape is causally relevant to the later appearance of another. We may grant that the designer might be able to favour macro-laws, and favour causally relevant macro-properties, without there being any laws or any causally relevant properties at the level of the micro-particles. But suppose that he is not allowed to work with lawless particles (they would represent an especially other-worldly feature); suppose that he has to establish micro-laws at the same time as macro-laws or that he has to take certain micro-laws as given. What happens then?

What must happen is that the designer organises things so that such micro-laws are satisfied—such micro-laws have their antecedents fulfilled—as mean that the macro-laws obtain. He must shape the micro-order, or he must select among the items to appear there, so as to rig things in favour of the desired macro-laws. But this then means that the program model is going to apply to the relation between the causally relevant properties at the two levels. If a certain aggregate shape is causally relevant to the later appearance of another shape, that is because its instantiation involves particles such that, by the micro-laws, they later move to a configuration sufficient to ensure the appearance of the other shape. The micro-properties will turn out to have causal relevance in a more basic sense than the aggregate properties.22

22 The fantasy may also explain why higher-level laws in the real world should be expected to be supervenient on the microphysical laws and the initial microphysical conditions. I think that if such supervenience is denied, that is probably because ‘laws’ are taken, not in the sense intended here—not in the sense of reliable regularities in rebus which our theories try to capture—but as the actual formulations of such regularities—formulations bounded by a variety of ceteris paribus clauses—that are found in the different sciences. If the ceteris paribus clauses for laws at different levels involve different idealisations—if they take us to different idealised versions of the actual world—then the idealised version of the actual world where the microphysical laws apply faithfully may not be a world at which some of the higher-level laws apply in that way. Thus there is at least one sense
The lesson of the fantasy is that the more basic causal status of the micro-properties envisaged does not suggest, for a moment, that the causal relevance of the aggregate shapes is suspect. While the micro-properties may be causally more basic in one sense, the aggregate properties are causally more important in another: they are the properties whose causal impact the designer is primarily concerned to establish. There can be no question of the causal relevance at the aggregate level being compromised by the applicability of the program model to the two levels in the example. And, by analogy, there should be no suggestion that the model introduces such a compromise in the real world.

We have been considering Macdonald's epiphenomenalist charge, that the program model makes higher-level properties causally impotent: only microphysical properties come out as potent in the proper sense. I have argued that the program model is consistent with there being no bottom level of causal relevance, and no causally efficacious properties of any kind, and that if it allows a bottom level—if it allows efficacious properties in sense 1—it still need not compromise the causal relevance of higher-level properties. Macdonald's charge has bite, I think, only under the assumption that not only is there a last level of causal relevance, but at this level properties exercise the irreducible, substantive sort of efficacy required by sense 2: they carry a punch—an oomph or bif—that is lacking in higher-level properties.

The point will be more or less obvious. Under the irreducibilist picture higher-level properties may enjoy the causal relevance of probabilifying the realisation of suitably productive properties. But they are not themselves productive of anything. The higher-level properties do not exemplify, as it were, the best candidate for what can be meant by causal power or potency. Microphysical

in which supervenience may fail. I am grateful to Peter Menzies for discussion on this point.

23 Intuitively, and roughly, there are two sorts of factors that may be taken as causes: controls and motors. Controls are entities such that by changing them you can, in principle, change the effect-factors; motors are forces which allegedly serve to produce the effect-factors in some unique sense of production. Causally efficacious properties in sense 1 are just basic controls; causally efficacious properties in sense 2 are motors. If there are motors to be found in the causal architecture of the world, then they will naturally represent the ultimate in causal power; controls will look second-grade, by comparison.
properties have a sort of causal relevance—an efficacy—lacking in all other properties.

But even if the epiphenomenalist charge goes through under the irreducibilist assumption, I do not think that it can be laid against the program model as such. For the picture of basic efficacious properties and causally second-rate, higher-level properties is an image cast by the irreducibilist assumption that there are causally efficacious properties in the substantive sense. It is not an image associated in particular with the program model. On the contrary, the program model has the virtue of showing how, even under this picture, the higher-level inefficacious properties can be still, at least, causally relevant; they can still program for certain results, by probabilifying the realisation of suitably efficacious properties.

One final question. Suppose you adopt the irreducibilist picture of bottom-level, causally efficacious properties and you endorse the program account of how higher-level properties manage to be causally relevant. Imagine that you still find yourself uneasy at the idea that certain higher-level properties are not efficacious in the sense in which the microphysical properties are so. Is there any move you can make to dissolve this anxiety? There is.

You can take a path to which Macdonald’s claims point. You can decide that causal efficacy is to be understood the easy way rather than the substantive or the formal way. You can cast the supervening properties which you take to be causally efficacious, and not just to be causally relevant in the programmatic way, as properties that are co-instantiated with efficacious physical properties. You can see them as properties whose instances are as potent as anything could ever be. I mention this move for the record and in order to point out the room for at least a partial reconciliation with Macdonald. But I must mention, in conclusion, two reasons why it may not be found reassuring.

The first is that if efficacy is to be understood in this fashion, then the notion is somewhat downgraded. Consider the change of shape that is causally relevant to the falling of the putty. This move will enable us to say that not only is that property causally relevant, it is also causally efficacious: it is co-instantiated with the microphysical property involving the arrangement of parts. So far, then, so good. But notice that the move will also force us to represent the change of volume as causally efficacious in producing
the fall. For, by parity of reasoning, this property will be co-instantiated too with the microphysical property. And so the move that is designed to give us comfort renders causal efficacy in such a way that causally irrelevant properties—properties that are intuitively irrelevant, and irrelevant under the program model—can count as causally efficacious.24

The second reason for wondering about how much comfort this move can give relates to difficult issues about the singularity of causation. The outright singularist about causality thinks that something which is the instance of a number of properties can be efficacious in a way that does not involve one of the properties instantiated any more intimately than others.25 He thinks that the fact that a property-instance is efficacious is a primitive fact, a fact that holds independently of what kind of instance it is. Under outright singularism, the move we are discussing would have us think of the supervening property as being causally efficacious in just the way in which the microphysical property is efficacious; both are efficacious through being instantiated in an efficacious instance. Thus, if we accept such singularism, and if we ignore the downgrading problem just mentioned, the move might give us a good degree of comfort.

But outright singularism is a difficult position to maintain, for it is plausible to think—certainly, it is widely assumed—that a

24 Macdonald mentions this sort of case, and suggests that there is a lesson to be learned from it when he says: 'the heat of the coal, not its redness, causes the burn'. 'The Nature of Naturalism', section 3. It would be nice if he could be persuaded that the lesson is this: for a property to count as 'efficacious' in a suitable, richer sense, it must be causally relevant, satisfying the sorts of conditions identified in the program model, as well as being co-instantiated with an efficacious physical property (as well as being efficacious in just 'the easy way'); the property must be like the change of shape rather than the change of volume. Sometimes Macdonald writes as if that suggestion ought to appeal: for example, in section 1 of his paper, when he discusses the computer case from which Jackson and I took the metaphor of programming. He writes there as if he endorsed something close to the program idea, even using the sort of terminology favoured by Jackson and me: ‘the computers are designed so that the instantiation of the program properties ensures that the lower order (hardware) properties produce effects which realise the intended higher order properties’. But when Macdonald mentions the redness of the coal case, the lesson he points up is a different and, to my eye, irrelevant one: that under the co-instantiation model, which supervening properties are efficacious will depend on the empirical matter of which physical properties are efficacious.

property-instance is going to be efficacious in virtue of the kind of thing it is, and not as a primitive fact; otherwise it must come as an empirical surprise that causality conforms to pattern and law. If the naturalist rejects outright singularity, however, then he is going to think that property-instances are efficacious in virtue of their microphysical character. And, thinking this, he will hold that the microphysical properties retain a sort of precedence over the higher-level properties with which they are co-instantiated. Thus, the move we are discussing will not save higher-level properties from the second-rate character which the irreducibilist assumption about causal efficacy foists upon them.26

26 I am grateful for comments received from David Braddon-Mitchell, Frank Jackson, Peter Menzies, Huw Price and Ian Ravenscroft.