

CRITICAL NOTICE

Mind and Cosmos

DAVID YATES

The central premises of Nagel's argument against what he terms 'the materialist neo-Darwinian conception of nature' (hereafter 'neo-Darwinism') are the following: (1) Remarkable features of the cosmos such as consciousness, cognition and value are *intelligible* to us; (2) such phenomena are not materialistically reducible; (3) only phenomena that are materialistically reducible are intelligible within the framework of neo-Darwinism.¹ In Chapter 2, 'Antireductionism and the Natural Order', Nagel argues for (1) and offers suggestive remarks on what 'intelligible' means, without ever approaching a definition. The remaining three chapters argue in turn that consciousness, cognition and value are not reducible, provide arguments as to why this renders them unintelligible for neo-Darwinists, and develop a sketch of an alternative way of understanding them. As Nagel puts it:

The essential character of such an understanding would be to explain the appearance of life, consciousness, reason, and knowledge neither as accidental side effects of the physical laws of nature nor as the result of intentional intervention in nature from without but as an unsurprising if not inevitable consequence of the order that governs the natural world from within. That order would have to include physical law, but if life is not just a physical phenomenon, the origin and evolution of life and mind will not be explainable by physics and chemistry alone. (32–3)

The explicitly secular alternative that Nagel considers is the addition of *teleological laws* to the non-teleological laws of neo-Darwinism, whose function is to render the emergence of the remarkable phenomena intelligible.

First, it should be noted that Nagel includes under the umbrella of reductionism positions whose proponents take them to be non-reductive, such as functionalism in the philosophy of mind (35–42). He clearly has in mind here a form of ontological reduction according to which the remarkable phenomena are, in some sense, nothing over and above material goings-on, so targeting supervenience forms of physicalism as well as more traditionally reductive alternatives such as the identity theory. For reasons of space I shall not discuss Nagel's arguments for anti-reductionism about consciousness, cognition and value. Suffice it to say that of the three arguments, readers are likely to find his familiar arguments for anti-reductionism about consciousness by

1 *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature is Almost Certainly False*. By Thomas Nagel. Oxford University Press, 2012. x + 130 pp. £15.99.

far the most persuasive. I shall focus on why Nagel thinks irreducible phenomena are unintelligible for neo-Darwinists.

Working this out is no easy task, not least because Nagel never really says what neo-Darwinism is supposed to be. Clearly, Nagel intends this conception to include more than just evolutionary biology; if not, more than just Nagel's remarkable phenomena will be unintelligible on it. Let us take it then that he means natural science, broadly construed, so as to include physics, chemistry, biology, genetics, natural selection and the rest. His aim, then, is to argue that the natural sciences, taken together, cannot render irreducible phenomena intelligible. Granting that they must be intelligible (of which more presently), what follows? If, as Nagel supposes, neo-Darwinism includes a commitment to materialism, it will be falsified by the mere *existence* of ontologically irreducible properties, so it is initially puzzling why Nagel would go about arguing that neo-Darwinism is false by arguing that it cannot render immaterial properties *intelligible*. Tellingly, Nagel only ever explicitly argues that the neo-Darwinian framework is *incomplete*. The key issue is what it would take to complete it, and Nagel's view is that only with the addition of teleological laws will natural science be able to supply adequate explanations of the target phenomena.

Grant Nagel that consciousness is irreducible (the form of the central argument is the same, once irreducibility is granted, for cognition and value), and that certain types of brain function give rise to certain conscious properties. Nagel suggests that explaining irreducible consciousness involves a two-fold task: (i) a constitutive psychophysical theory connecting various types of brain function to various conscious properties in a systematic way, and (ii) an historical evolutionary theory that explains how we got to have brains that function in the ways from which our conscious properties emerge. Nagel expresses doubts about the possibility of (i), but does not have a knock-down argument, and I shall not discuss his objections here.

Throughout the book, Nagel appears to deny that neo-Darwinist historical explanations could render *any* properties of living organisms intelligible, not to mention the remarkable phenomena he takes to be irreducible. For instance:

The explanation by standard evolutionary theory of the purely physical characteristics of organisms is hard enough even if one disregards consciousness. [...]The theory of natural selection, if it is to rely only on the operation of physical law, has to postulate that there is a purely physical explanation of why it is not unlikely that accidental mutations in the genetic material have generated the range of variation in viable phenotypes needed to permit natural selection to produce the evolutionary history that has actually occurred on earth over the past three billion years. (48)

He goes on to claim that 'standard evolutionary theory', by which I again assume he means natural science, broadly construed, cannot render any of the purely physical characteristics of living things *probable*, and so fails to explain even them. Here, as elsewhere, Nagel apparently endorses the widely discredited view that to explain something – to render it intelligible – is to render its occurrence more probable than not, given antecedent conditions. On this view, that X is a heavy smoker does not render intelligible that X dies of lung cancer, because the probability of X so dying conditionally on X being a heavy smoker is significantly less than 0.5. Such cases are no small part of the reason why this theory of explanation is less than widely endorsed among philosophers of science.

Nagel also tacitly assumes that ‘accidental mutations in the genetic material’ are not deterministically caused by prior conditions according to physical laws, because if they were, we could in principle predict such mutations if we knew the physical mechanisms that underpin them. Nagel must be assuming that our world is not deterministic. Fair enough, but if to explain is to render probable, and genetic mutation is *indeterministic*, then it is hardly surprising that the cumulative effects of three billion years of such mutations come out unintelligible.

It is unfortunate that Nagel makes these implausible claims linking explanation and probability, because his most interesting arguments do not require them. Nagel explains that:

... a conjunctive explanation, going from A to B and B to C, can explain C only if there is some further, internal relation between the way A explains B and the way B explains C.[...]It isn’t enough that C should be the consequence, even the necessary consequence, of B, which is explained by A. There must be something about A itself that makes C a likely consequence. (52)

Suppose we have a constitutive theory of consciousness according to which certain types of brain function systematically necessitate conscious properties, and an evolutionary theory that successfully explains how we got to have brains that function that way. To make the point vivid, suppose the historical theory is deterministic, and that the physical facts about our evolutionary history necessitate our having brains that function that way. The passage just quoted implies that Nagel thinks consciousness will *still* be unintelligible, unless its emergence is explained by *the historical theory alone*. Of course this is impossible if consciousness is irreducible, but that, as I see it, is Nagel’s point. The principles of the constitutive theory, according to which specific brain functions give rise to specific conscious properties, play no role in the evolution of brains capable of performing those functions; similarly, the dynamic principles of the historical theory play no role in the emergence of consciousness from brain function. As is clear from his discussion of panpsychism as a constitutive theory, Nagel thinks that ‘conjunctive explanations’ of the form outlined above work only if the historical and constitutive theories are based on the same set of unified principles, but the irreducibility of consciousness rules this out (see in particular 54–65).

It seems then that Nagel holds rendering probable to be at most a *necessary* condition on rendering intelligible. If the above interpretation of his argument is correct, Nagel need not endorse this condition, because on that interpretation the argument does not depend on denying that neo-Darwinian theories supplemented with the appropriate constitutive theories can in principle meet it; nor does it depend on denying the possibility of constitutive theories. The gap in our understanding that Nagel sees is this: if the physical laws that govern the dynamic evolution of our world have nothing to do with the emergence of consciousness, the fact that those laws produced brains that meet the conditions for consciousness imposed by the constitutive theory is a *coincidence*.

How are teleological laws supposed to help? Nagel’s thought, it seems, should now be that such laws make the coincidence go away because consciousness is somehow encoded into the dynamic laws – those laws are the way they are *in order that* they someday produce brains that give rise to consciousness, and what would otherwise

seem to be a coincidence was supposed to happen all along. Frustratingly, when Nagel discusses natural teleology, he appears to suggest that rendering probable is both necessary *and sufficient* for intelligibility, which implies that the problem with neo-Darwinian conjunctive explanations is that they do not raise the probability of consciousness, cognition and value *high enough*:

Natural teleology would require two things. First, that the nonteleological and timeless laws of physics...are not fully deterministic.[...]Second, among...possible futures there will be some that are more eligible than others as possible steps on the way to the formation of...the kinds of replicating systems characteristic of life. The existence of teleology requires that successor states in this subset have a significantly higher probability than is entailed by the laws of physics alone – simply because they are on the path toward a certain outcome. (92–3)

Nagel's characterization of the problem of conjunctive explanation suggests that the source of the problem is the theoretical disunity of the historical and constitutive principles required to explain remarkable, irreducible properties. On this account, the problem is that it is a coincidence that the dynamic laws produced a world in which the constitutive principles are operative. But if this is what Nagel has in mind, why make such controversial claims about the theoretical adequacy of neo-Darwinism with respect to the evolution of purely physical characteristics? His speculative remarks about natural teleology as a solution to the problem, on the other hand, suggest that the problem is indeed that conjunctive explanations where the historical principles are neo-Darwinian do not work because to explain is to render probable, and non-teleological dynamic laws cannot do that for the physical properties of living things, never mind their remarkable properties. But if this is what Nagel has in mind, why spend so much time arguing that the remarkable properties are irreducible?

A charitable reading of Nagel is that he thinks *both* that neo-Darwinism does not render the evolution of purely physical characteristics intelligible, and that even if it did, the irreducibility of consciousness, cognition and value preclude its rendering their emergence intelligible. But now we have two distinct notions of intelligibility in play. Let us say that a theory T renders the emergence of some remarkable phenomenon x intelligible₁ iff T renders x 's emergence probable; that is, $P(x/T) > 0.5$. Let us say that T renders x 's emergence intelligible₂ iff x 's emergence is not a coincidence given T . These two notions of intelligibility are orthogonal. T might render x 's emergence intelligible₁ without rendering it intelligible₂, since raising the probability of x 's emergence above 0.5 does not require the kind of theoretical unity necessary, according to Nagel, to render its emergence non-coincidental. Similarly, T might render x 's emergence intelligible₂ without rendering it intelligible₁, since rendering x 's emergence non-coincidental does not require that its emergence be rendered more probable than not. According to Nagel, neo-Darwinism fails to render consciousness intelligible₁ in the sense that it does not render the brain states from which it emerges probable. But even if neo-Darwinism did render consciousness intelligible₁, it fails to render it intelligible₂ in the sense that the historical and constitutive principles required to render it intelligible₁ are unrelated, so it is a coincidence, given only the historical theory, that we arrived at a point where the constitutive principles apply. Since the role of

teleological laws is to fill the gaps left by neo-Darwinism, the two notions of intelligibility to which Nagel appeals suggest a dual role for those laws: raising the probability of living things to greater than 0.5, and unifying the historical and constitutive principles involved in conjunctive explanations of their remarkable properties.

Perhaps the most serious problem for Nagel's overall approach is his commitment to the intelligibility of the cosmos. Early in Chapter 2, Nagel tells us that '[s]cience is driven by the assumption that the world is intelligible', which suggests that he expects the assumption of intelligibility to be shared by proponents of the neo-Darwinian conception of nature he hopes to overturn (16). Perhaps there is some sense of 'intelligible' in which this proclamation is true, but it is not either of the senses upon which Nagel relies in his argument. Nagel argues that teleological laws are required to render intelligible *only* remarkable phenomena such as consciousness, cognition and life. It follows that much – indeed, most – of the natural world is governed by non-teleological laws alone. It is quite pleasant outside today; it is not cold or raining, and there is a little sunshine, but there is nothing *remarkable* about today's weather that would require us to posit teleological laws in order to explain it. But if, as Nagel assumes, non-teleological laws are indeterministic, and science requires the world to be intelligible₁, then the unremarkable aspects of our world typically will not be amenable to scientific investigation. Unless Nagel wishes to say that the laws of physics that govern the weather have always been the way they are so that today's weather would be quite pleasant, then it seems meteorology is not scientific. If science requires intelligibility₁ and the laws of nature are indeterministic, then either teleology is ubiquitous, or science is rare.

The supposition that science requires the intelligibility₂ of the cosmos is even less plausible, because this amounts to the claim that science is only possible if there are no coincidences, and Nagel of course does not believe that. In fact, he explicitly claims that it is only remarkable phenomena that must be intelligible₂, granting that many unremarkable phenomena such as today's weather are indeed coincidences. But according to which objective definition of 'remarkable' does consciousness count as something remarkable, whereas a pleasant day in the countryside does not? Nagel does not say, but it is difficult to avoid thinking that whether or not some phenomenon is remarkable depends as much upon our attitudes towards it as upon the phenomenon itself. A guiding idea of the book seems to be that phenomena of such significance as consciousness simply could not have evolved without a *prior purpose*. But if this significance is subjective, then whether or not the intelligibility of a phenomenon requires it to have such a purpose is likewise subjective.

An exploration of the thesis that our standards of intelligibility increase with the cognitive significance of the explananda would have been most welcome, but that thesis would hardly justify the claim that neo-Darwinism is incomplete, let alone false, and in any case it is not what we get here. The book is not without its merits – it is certainly thought-provoking, and the idea of natural teleology is of sufficient independent interest to merit a more sophisticated treatment than Nagel gives it here. Unfortunately, the book is let down throughout by a lack of clarity and precision, and is as likely to frustrate as it is to inspire.

*Corpus Christ College
Merton Street, Oxford OX1 4JF
david.yates@philosophy.ox.ac.uk*