# Nagel, Panpsychism and Realism

**NICHOLAS UNWIN**

## Abstract

Thomas Nagel argues that panpsychism (that everything has mental qualities) follows from four premises, including what he calls 'Realism'. This paper argues that he misidentifies the sort of realism that he should be concerned with, and that a useful parallel can be drawn with certain nonrealist theories of ethics, notably expressivism. An expressivist or projectivist account of conscious states is argued to be more plausible than it looks, and that adopting such a position would successfully block the argument for panpsychism. Comparisons are also drawn with David Chalmers's position.

In his celebrated paper of that name, Thomas Nagel argues that panpsychism, namely the doctrine that the basic physical constituents of the universe have mental (or 'protomental') qualities regardless of whether they form parts of living beings, follows from four plausible premises:

1. **Material Composition**

   Any living organism, including a human being, is a complex material system. … No constituents besides matter are needed.

2. **Nonreductionism**

   Ordinary mental states … are not physical properties of the organism … and they are not implied by physical properties alone.

3. **Realism**

   [Mental states] are [real] properties of the organism.

4. **Nonemergence**

> All [nonrelational] properties of a complex system … derive from the properties of its constituents and their effects on each other when so combined.[1]

The derivation runs as follows:

> If the mental properties of an organism are not implied by any physical properties but must derive from properties of the organism's constituents, then those constituents must have nonphysical properties from which the appearance of mental properties follows when the combination is of the right kind. Since any matter can compose an organism, all matter must have these properties. And since the same matter can be made into different types of organisms with different types of mental life (of which we have encountered only a tiny sample), it must have properties that imply the appearance of different mental phenomena when the matter is combined in different ways. This would amount to a kind of mental chemistry.[2]

Panpsychism is not a plausible looking doctrine, so either the derivation fails or (more probably) at least one of the premises must be false. Much of Nagel's article consists of examining carefully each premise in turn.

An oddity of the article is that a good deal of attention is directed at the Realism premise. Nagel retains some considerable misgivings about its truth, and yet as it stands it seems very minimal. Of course, mental states are really properties of the organism, we might say: how could they fail to be? Nagel never explains very clearly what the alternative might be, though he discusses Wittgenstein at some length. This paper will argue that a useful comparison can be drawn with moral properties. Antirealist theories of ethics, notably expressivism, deny that moral properties are ordinary factual properties, and a different account needs to be given of what we

[1] Nagel, Thomas (1979), 'Panpsychism', in *Mortal Questions*. Cambridge: Cambridge University Press: 181–95 at 181–2.
[2] Ibid, 182.

are doing when we make a moral judgement. Specifically, moral judgements do not express genuine beliefs (or not wholly so), but rather noncognitive attitudes of various kinds, such as feelings and desires. Such a comparison may seem surprising, but there are a number of points in its favour.

Firstly, it seems consonant with Nagel's own understanding of what it is to be conscious: for *x* to be conscious, there must be something which it is like to be *x* (and vice versa). That is what consciousness *is*. Yet the property of there being something which it is like to be that thing is not straightforwardly graspable and is something of a mouthful. What is easier to grasp is the idea of imagining what it is like to be *x*. Here one is projecting a kind of empathic attitude. But it is the attitude rather than the projected property that is uppermost, and this suggests an expressivist analysis. If I imagine what it is like to be *x*, then I am not making a straightforward cognitive judgement; so a connection with noncognitivist theories of ethics becomes apparent.

Other thinkers have also presented views which suggest a similar conclusion. There is, for example, the famous remark of Wittgenstein's: 'My attitude towards him is an attitude towards a soul; I am not of the opinion that he has a soul.'[3] If I see someone in pain, I do not merely see, in a detached sort of way, a human body exhibiting pain-behaviour. On the contrary, I interact, empathize and so forth, and these attitudes form a significant part of my recognition that this person is in pain. This is not quite the same as saying that to judge that a person is in pain just *is* to express certain noncognitive (as well, of course, as cognitive) attitudes of a certain kind; but it gets close. At any rate, the distinction between opinions and attitudes is clearly expressivist in tone.[4] The central point is that in making a judgement about other minds, I am engaged in a kind of empathic projection, and projections also underpin expressivist theories of ethics. A related point is that, even if mental reality is ultimately nothing over and above physical reality, judgements about minds seem to involve more than just judgements about physical

[3] Ludwig Wittgenstein (1958), *Philosophical Investigations*. Oxford: Blackwell. Iii.
[4] Simon Blackburn considers Wittgenstein's dictum here, though does not explore the expressivist implications in any detail. See his 'Moral Realism', in *Essays in Quasi-Realism*. Oxford: Oxford University Press, 1993: 111–29 at 112–3.

things and their physical properties. This applies as much to other minds as to oneself. Mental judgements just do not look like physical judgements, and this can lead us to think that there must be more to the mental than just the physical. But it may be that the gap is simply in the judgement, not the reality. This precludes dualism, but it also precludes a straightforward physicalism. If mental states were physical states, then to attribute a mental state to someone would simply be to attribute to him that physical state (perhaps described differently). The projections would drop out of the system as irrelevant, and they do not seem to do so.

Moral properties show some similar features. They do not seem to be reducible to physical properties, and we have a gap between fact and value: factual statements do not entail value judgements. This echoes Nagel's nonreductionism premise. But we nevertheless seem to have a supervenience connection which rules out anything analogous to dualism. That is to say, it is incoherent to judge two situations to be factually alike but morally different; and similarly it seems impossible to have two people who are physically indistinguishable but mentally different The possibility of 'zombies', or unconscious physical duplicates of ourselves, is often alleged, and this complicates the picture; but Nagel himself accepts mind-brain supervenience along with nonreductionism, so the question of whether zombies are possible can be shelved as far as the discussion of Nagel's own position is concerned.

The brain somehow generates consciousness, but we do not understand how it does it or why it should do it at all. This is what David Chalmers calls the 'hard problem of consciousness'. It would be a considerable advance to be able to see even what the basic shape of the solution might be, regardless of details. A step forward, I think, is to see that we do not seem to have a hard problem of values in quite the same way. There ought to be such a problem given that there is a fact/value gap, for if the ordinary facts of the world do not entail values, then how do the values come about? Must we not cite what J.L. Mackie calls 'queer' entities (objective values), just as a dualist needs to cite Cartesian minds, or at the very least nonphysical properties or

mechanisms of some kind which are equally queer?[5] However, an expressivist theory yields a metaphysically straightforward alternative answer. Value judgements express noncognitive mental states, not genuine beliefs, and it is nothing more mysterious than this which ensures that there is no inference from fact to value since there is, in Hume's words, a 'real distinction' between beliefs and 'passions'. If there is also a gap between our beliefs about physical things and their constituents, on the one hand, and the empathic judgements that we make when we address other minds, then perhaps we can also see why consciousness is irreducible to the physical, but in a way that is relatively harmless. We don't have any mysterious extra qualities in the fabric of the world, just an additional set of judgements of a different kind.

It is helpful to compare different theories of mind with different theories of values. Thus the property dualism of Chalmers, a kind of mental realism, corresponds to nonnaturalist moral realism, for example G.E. Moore's intuitionism. Thus, on the one hand, we have real conscious properties which are irreducible to physical properties, and on the other hand, we have objective moral properties which are not reducible to what Moore calls 'natural properties' (ordinarily observable physical and mental properties). The only significant difference is that Chalmers insists on the metaphysical possibility of zombies, which means rejecting supervenience. Almost all moral philosophers, by contrast, think that values supervene on facts. Moore, for example, accepted this (he was probably the first to formulate the notion though he did not use the term). However, it is hard to see how to justify supervenience given nonnaturalist realism (that supervenience should hold seems to be something of a mystery here), so the analogy remains fairly good. By contrast, naturalist moral realism corresponds to physicalism. Moral properties just are natural properties of a certain sort, and mental properties just are ordinary physical properties.[6] Nonrealist moral theories, such as expressivism for example, however, do not have any obvious counterpart in the philosophy of mind. There are suggestions to be found in

[5] J.L. Mackie (1990), *Ethics: Inventing Right and Wrong*. London: Penguin.
[6] Functionalist or multiple realizability theories, such as Putnam's, likewise correspond roughly to 'Cornell realism' in ethics.

Wittgenstein, as mentioned, and possibly also Dennett who, at least at one time, insisted that mental judgements are projections from an 'intentional stance' and do not reflect ultimate patterns in nature.[7] But mostly the possibility of a projectivist solution to the mind–body problem is not considered.[8] This, at least, is true in the Western tradition.[9]

Moore's nonnatural qualities are indeed mysterious, not least of all because they too would have to inhere in the components of actions as well as the actions as a whole. This follows because if the constituents have only natural qualities, then the whole action can only have qualities that are composed of natural qualities. The nonnatural qualities cannot just emerge from nothing, any more than ordinary mental qualities can: that is the essence of Nagel's nonemergence premise. So, just as panpsychism lands us with 'protomental qualities', qualities that fall a long way short of full consciousness but which form the building blocks of all kinds of conscious experience, human or otherwise, moral nonnaturalist realism lands us with what we may call 'protomoral qualities', qualities that give rise to full-blown moral qualities by means of a sort of 'moral chemistry'. Such a view is seriously odd, though perhaps not so much so that it can be ruled out decisively. But it would clearly be better if we could do without 'protomoral qualities', and the crucial premise that leads us to them is realism. This is, as far as I know, a new objection to moral nonnaturalist realism. It also suggests how a similar rejection of realism in this sense could block the need for protomental qualities and panpsychism.

It may be protested that the very notion of a protomoral quality is too bizarre to be taken seriously in the first place, and that we do not have a good analogy here. But this is not so obvious that it can be taken immediately for granted. Moral situations tend to be complex, and may involve several agents, and each individual action may have several significantly independent

[7] Daniel C. Dennett (1989), *The Intentional Stance*. Cambridge MA: MIT Press.
[8] Though see Georges Rey (1995) (1995), 'Towards a Projectivist Account of Conscious Experience', in Thomas Metzinger (ed.), *Conscious Experience*. Paderborn: Ferdinand Schöningh: 123–44. Rey does not develop the idea in the same way that I do, however.
[9] Ram-Prasad Chakravarthi informs me that projectivist ideas may be found in Mahayana Buddhism and Advaita Vedanta.

components. In asking what makes the whole situation the way it morally is, we need to look at the morally relevant features of the components and how they form a whole. The distribution of moral qualities thus exhibits a good deal of internal structure, a structure which plays a central role in the whole picture. What is far more contentious is the thought that we can take this further, and reduce these components into even smaller components, and so on down to basic physical constituents. The reason for rejecting this view is, surely, that these ultimate protomoral qualities do not seem to play any significant role in our moral thinking. They act simply as metaphysical baggage, and do not fulfil any visible moral purpose.

Still, we need to explain how alternative theories can avoid such things. The virtue of nonrealist moral theories is that attention focuses on the attitudes expressed in the making of a judgement. Now, here too there can be internal structure. In the morally complex situation discussed above, I may have a complex raft of attitudes that exhibit an internal structure. Thus given a morally relevant situation X, there will be subsituations $X_1$, ..., $X_n$ which constitute it. Likewise, my attitude A towards X will be divisible into attitudes $A_1$ towards $X_1$, $A_2$ towards $X_2$, ..., $A_n$ towards $X_n$. But, crucially, there is no obvious urge to stretch things to the limit and to postulate microattitudes (what remains when the division of A into $A_1$, ..., $A_n$ is further divided and stretched to the limit) the expression of which might be thought to constitute judgements about protomoral qualities. In simple versions of moral nonrealism, such as emotivism, the central attitudes are ones of approval and disapproval, and these are understood to be simple. I can approve and disapprove components of a complex action, of course, but the act of approval itself does not have component subattitudes. Moreover, this does not seem like a disadvantage. More sophisticated expressivist theories, such as those presented by Simon Blackburn and Allan Gibbard behave similarly. Of course, there can be complexity within attitudes, but it does not have to exhibit the kind of features that might have led us to postulate protomoral qualities.

Nagel's own reasons for having misgivings about Realism are rather different:

> For Realism as I have defined it to be true, physical organisms must have subjective properties. What seems unacceptable about this is that the organism does not have a point of view: the person or creature does. ... When a mouse is frightened it does not seem to me that a small material object is frightened.[10]

So although it is true in a sense that the mouse is frightened, it is not really the (organic) mouse itself that has the property of being frightened: in fact, nothing 'out there' has that property. At least, it does not appear that way. Nagel concedes that this intuition does not seem to go anywhere, and he eventually gives up his resistance to Realism. But there does seem to be something perverse about seeing the frightened mouse as a frightened small material object. Why is this? It is, I suggest, because the sort of attitude you have when you judge *x* to be frightened is very different from the sort of attitude you have when you judge *x* to have purely material qualities. In short, it is the attitude which is distinctive. But this is to invoke rather different considerations to those to which Nagel appeals.

The word 'realism' has many philosophical meanings, of course, but even with Nagel there is more than one sense. On the one hand, there is the capitalized, slightly odd sense that forms his third premise, that mental properties are genuinely properties of the organism. But there is the more recognizable sense which is involved when he insists that phenomenal consciousness has a reality and definite nature that transcends our ability to understand it. Thus, there is definitely something which it is like to be a bat, even though human beings are cognitively constrained in such a way that they can barely glimpse what this is (the bat point of view is different from the human point of view).[11] There is thus a strong anti-verificationism at work here, a confidence that there are many genuine truths of which we shall forever be unaware. The view that I am suggesting, by contrast, is rather different from both these senses.

[10] Nagel 1979: 189.

[11] Physical properties, by contrast, can be grasped by intelligent creatures of any species. This is Nagel's main reason for supposing that mental properties are irreducible to physical properties.

Yet there is a connection. If bat-consciousness has a reality of its own, something quite independent of my understanding of it, then judging that a bat is conscious must be nothing more than simply judging that the bat has this objective property. Any expressive projection of my own attitude towards it will thereby drop out of the picture as irrelevant, as we noted earlier. So if, as I suggest, this attitude is not irrelevant, the bat's consciousness cannot be given a realist treatment, albeit not in quite any of the senses of 'realist' that Nagel uses.

This needs qualification. We had better not be committed to saying that the bat is conscious only because we judge it in a certain sort of way: that is clearly wrong. However, an expressivist or projectivist account does not commit us to saying that. For example, Simon Blackburn, in a well known passage, insists that

> [i]f most of us come to taste phenol-thio-urea as bitter, then that is *what it is* for the stuff to become bitter. If most of us come to find wanton violence admirable, that is not what it is for wanton violence to become admirable: it is what it is for most of us to deteriorate in a familiar and fearful way.[12]

It is just wrong to suppose that wanton violence is wrong because we disapprove of it, if that implies that, had we not disapproved of it, it would not have been wrong. There is a distinction, which Blackburn emphasizes, between projectivism and response-dependence theories. With the latter, the quality (for example, a secondary quality) is actually constituted by our being disposed to judge things in a certain way, and this gives such qualities an irreducible subjectivity. Projectivism is different, for we project also into hypothetical situations where our attitudes are different from the way that they actually are, and in such a way as to yield the contrast expressed in the quoted passage.

[12] Simon Blackburn (1981), 'Reply: Rule-Following and Moral Realism', in Steven Holtzman and Christopher Leich (eds), *Wittgenstein: To Follow a Rule*. London: Routledge & Kegan Paul: 163–87 at 174–5.

My view is unusual in so far as it supposes that secondary qualities in the objects should be given a response-dependent treatment, but supposes that the perceptions themselves be given a projectivist analysis. Thus pigeons, like bats, have a type of phenomenal consciousness very different from ours. For example, they have a tetrachromatic (possibly pentachromatic) colour vision system which is vastly more sophisticated than anything to be found in mammals (its evolutionary history is far longer). We cannot imagine what it is like to see pigeon hues, but suppose that it is nevertheless like something, something that might even be phenomenally richer than anything we ever experience, at least in some aspects. And this something has a reality that exists regardless of the limits of what we can imagine. But in fact we would never be disposed to attribute colour perceptions to pigeons were it not for the fact that pigeons' eyes and brains resemble our own in numerous ways. Such attributions are thus not just idle speculation but are grounded in observable similarities to ourselves. The same would not be true if we were to attribute consciousness to pebbles, for example; so claiming that such entities could be conscious after all involves more than just a faith in verification-transcendent truth. Such claims are bizarre to the point of unintelligibility; and the reason, I claim, is that we lack the ability to adopt the appropriate sort of attitude towards pebbles. We can empathize with pigeons and bats up to a point, but cannot empathize with pebbles even up to a point.

This could be disputed, perhaps. Thus Ted Hughes (in *Pibroch*) writes:

> ... A pebble is imprisoned
> Like nothing in the Universe.
> Created for black sleep. Or growing
> Conscious of the sun's red spot occasionally,
> Then dreaming it is the foetus of God.
>
> Over the stone rushes the wind

Able to mingle with nothing,

Like the hearing of the blind stone itself.

Or turns, as if the stone's mind came feeling

A fantasy of directions.[13]

Yet this does not license a breach between what we judge to be conscious and what we can empathize with: rather, it just shows that the poetic imagination has a very wide empathic reach. It remains unintelligible for us to wonder whether *x* is conscious without wondering whether we can empathize with *x*. There therefore still needs to be limits to the sort of realism that Nagel advocates. We cannot really make much sense of the notion of pebble-consciousness, so should resist the thought that it just might nevertheless exist anyway albeit in some wholly unimaginable form. We must also remember that our ruminations about what it might be like to be a bat are more closely rooted in what we humans can make sense of than might be thought, and this sets limits to the type of realism, in the sense of acceptance of verification-transcendent truth, that is acceptable.

Our attitude towards pigeons is subtly different from our attitude towards bats, and this is because of their observable physical differences. This can be generalized to all living organisms sufficiently complex that we are prepared to attribute consciousness to them. But such attitudes need not be generalized to such an extent that they be attributable to the constituents of these organisms. We have no serious inclination to ask what it is like to be a pigeon's retina or a small part of a bat's brain, nor should we. Protomental properties thus never get a chance to enter the picture.

It may be protested that we are in danger of proving too much. Consciousness, or at least some sort of bare mentality, is widespread and extends a long way down the phylogenetic tree. A point which Chalmers emphasizes is that it seems odd to suppose that there should be a

[13] I am grateful to John Foster for drawing my attention to this passage.

discontinuity here, a point where consciousness just 'winks out'.[14] If there were, then something little short of miraculous would have to happen when we move upwards through the critical point and consciousness 'winks in' (William James made a similar point). He is also prepared to attribute consciousness (albeit of a massively rudimentary kind) to thermostats, on the grounds that they are information-processing systems (a subsection of *The Conscious Mind* is titled 'What is it like to be a thermostat?'). However, even if we go along with this, we should still balk at supposing that the mental properties of a complex system are constructed out of those of its constituents in any useful way. The nonemergence premise insists that we do this, of course, at least when applied to real intrinsic properties. Blocking the realism premise is therefore still a valuable exercise, and may indeed be the only effective way of preventing such excesses.

It may be protested that all this is to neglect a more fundamental objection to Nagel's view, namely concerning the nonemergence premise itself. It is often regarded as the weakest premise.[15] Nagel defends it on the grounds that we would otherwise be left with (what is rightly or wrongly usually called) a Humean theory of causation, one where connections arise for no metaphysical reason at all. But it may be felt that this is to fail to do justice to Hume's point that laws of nature are ultimately things that cannot be explained. We can explain some laws in terms of more basic ones, of course, for example Kepler's laws in terms of Newton's laws. But the fact that we have an inverse square law of gravitation, as opposed to an inverse fourth power law, is just a bare fact about the universe. There is no metaphysical impossibility in the alternative law, and in some possible worlds it will hold. If you want to know why our world instead exhibits the inverse square law, there is not much of an answer that can be given. We tend to think that the ultimate laws of nature will be laws of physics, but as Chalmers points out, there is no obvious reason why this should have to be so. Such laws might also include psychophysical laws, laws

[14] David J. Chalmers. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press, 293–301.
[15] See, for example, William Seager and Sean Allen-Hermanson, 'Panpsychism', *The Stanford Encyclopedia of Philosophy (Fall 2010 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2010/entries/panpsychism/>. §4.1.

that say that whenever a certain type of physical structure is realized, consciousness will be present, and they too are inexplicable. They just emerge *ex nihilo*.

However, it is one thing to allow that fundamental laws of nature are not metaphysically necessary truths; it is another to suppose that a complex system can have properties which fail to relate to properties of the constituent parts.[16] Chalmers himself is not averse to panpsychism, and he agrees that consciousness needs to be 'integrated into the natural order'.[17] This integration could no doubt be understood in many ways, but it presumably will include a harmony between physical and mental complexity, and a sensitivity to how things are constructed from their constituents. More generally, anyone who objects to the nonemergence premise might be challenged to provide another counterexample to it, and this may not be easy: outside quantum mechanics, wholes really are typically explicable in terms of their parts.[18]

We may still wonder whether my projectivist view is more plausible than Chalmers's. Conscious thermostats may not be everybody's cup of tea, and the whole business gets very close to panpsychism itself, but attempts to extend expressivist analyses to the mind–body problem are hardly intuitive either. However, there is a general problem about psychophysical laws which Chalmers's theory faces and my theory does not. This concerns the general shape of such laws. A law that says that whenever such-and-such physical features are to be found so is consciousness is not really satisfactory as a *fundamental* law. This is partly because there are going to be too many candidates for 'such-and-such', but also because consciousness is not a single uniform quality common to all and only conscious states. You cannot, for example, transform your zombie twin's functional states into the full-blown mental states that you yourself experience by just adding to them a single extra ingredient, 'Consciousness', unless this ingredient is very artificially contrived, and therefore quite out of place in any serious law of

[16] Galen Strawson (2006: 12–20), in his own defence of panpsychism, also emphasizes this point.
[17] Chalmers 1996, 298.
[18] This is not to deny that holistic explanations are often appropriate; for example, to understand the behaviour of a pound coin in circulation, we have to embed it into a wider institutional environment. This does not invoke emergent properties in any controversial sense, however.

nature. When a person sees green, for example, then she is in a unified state that cannot be decomposed into two separate parts: the purely functional part that her zombie twin also exhibits; and the state of pure consciousness, i.e. the state of there being something which it is like to be her. More generally, our phenomenology is too rich and diverse to be understood simply as a single, uniform there-is-something-which-it-is-like-to-be-*x* tacked onto a variety of zombified residues. We simply lack an account of how the pure consciousness and the zombified residue fit together to form a seamless whole. This difficulty of breaking conscious states into components makes it hard to see how fundamental laws of consciousness can be formulated.

A projectivist view need not search for such laws. Empathic projections yield a conception of third-person consciousness that does not require the sort of explanation that needs laws. Again, a comparison with the case of moral values is instructive, for we have no analogous problem here unless, perhaps, we adopt a nonnaturalist realism. If we are disposed to ask what it is like to be *x*, but not disposed to ask what it is like to be *y*, it is not because *x* has, and *y* lacks, some real intrinsic property called consciousness, or that we suppose this. The features which distinguish *x* from *y* are ordinary observable ones, not mysterious ones lurking behind the scenes. And the attitude of empathy, which underpins the thought that there is something which it is like to be *x*, is not a narrowly cognitive attitude, a belief that *x* has a certain factual property, but something which draws on a wider range of mental attributes.

To summarize: Nagel is right in thinking that his argument for panpsychism needs a realism premise, but he misidentifies the sort of realism that is at stake. His own conception of 'Realism' is too thin to make much of a difference, but the distinction between realist and nonrealist theories of ethics is more to the point. Nagel's own understanding of what it is to judge that *x* is conscious suggests a projectivist view of mental states, and this projectivism has several advantages when compared with, for example, Chalmers's view of consciousness and

what ultimately underpins it. And even if it is still held to be too implausible to be accepted, it remains true that we need a nontrivial premise that excludes it.[19]

Lancaster University

N.Unwin@lancaster.ac.uk

[19] An earlier version of this paper was given to the Philosophy Work in Progress seminar at Lancaster University, and I am grateful for many useful comments and suggestions.