

TONY CHADWICK ESSAY PRIZE 2003 WINNER

RESPONSIBLE BRAINS

by Peter Gibson

Philosophers who take a very reductive attitude to the mind seem to be the demons of our age, because they threaten moral anarchy if we cannot clearly distinguish humans from the rest of nature. This threat arises because reductive physicalism doesn't seem to allow room for free will, or pure reason, or personal identity, which are the foundations of most moral thinking. This essay offers a rapid outline of why reductive physicalism seems to be correct, then suggests that we should accept the consequent absence of free will, but then indicates a strategy whereby physicalists can retain a central place for morality.

The main opponents of reductive physicalism these days are a group of eminent thinkers, labelled by John Perry (2001:22) as 'neo-dualists', who try to give mind the special status which seems to be needed for a moral life, while remaining within a materialist framework. I will first try to sketch the reasons why the approaches of Kripke, Jackson, Nagel, Chalmers and Davidson fail.

1

Kripke's modal argument (1972:144-155) has its origin in Descartes' argument concerning necessity of mind and contingency of body (1641 :156). Kripke first clarified the concept of identity in terms of possible worlds. A true identity, such as 'Hesperus' being 'Phosphorus' (they are both the same planet, Venus), will hold in all possible worlds, and so is necessary. Unfortunately the early physicalists, such as J.J.C. Smart (1959), asserted that the identity of mind and brain was contingent. Kripke argued that since the physicalists claimed contingent identity, and he had shown that there is no such thing, physicalism is false.

But Smart was guilty of a confusion of ontology and epistemology. The identity of mind and brain is held to be contingent because it is a matter of scientific investigation, but this is epistemology. The mind and the brain are either identical or not, just as every number above two may turn out to be the sum of two prime numbers, or not. The facts will be necessary, despite the contingency of whether we discover them. Strangely, instead of simply pointing out this error, Kripke chose to argue that the identity theory is wrong, by resurrecting Descartes' claim. Sensible physicalists (e.g. Perry 2001) now assert a necessary (de re) identity of mind and brain, and hence the modal argument no longer has an opponent.

2

Frank Jackson's knowledge argument (1986) is exemplified in the story of 'Mary', who has all the theoretical knowledge of colour, and then acquires 'new knowledge' when she first experiences it. This implies that facts about experience are not facts about brains, and so they cannot be identified with one another.

The argument must claim that Mary acquires a new type of knowledge, which seems to require a new type of fact about reality. Simply having a novel experience is not enough to justify an ontological claim about the mind, otherwise the first person to experience magnetism could have claimed to have a different order of consciousness from the other people around them.

Jackson focuses on the advent of colour in Mary's world. This is certainly a novelty, and the example invites us to imagine her stunned surprise, and hence feel colour perception as almost miraculous. However, it seems wrong to say that she experiences a new type of knowledge. She already has sight, and has been experiencing a range of grey tints for years, so her sight of

redness is actually only an accelerated expansion of her current repertoire. This seems to be no more new knowledge than if I expanded my repertoire by seeing a new shade of amber, or hearing the sound of a new musical instrument.

Even if we pushed the example further, and raised her in total darkness (listening to audiotapes), it does not seem that the advent of vision is a new type of knowledge. She always knew where her chair was, but she now has a new way of detecting it. If we accept that knowledge is (roughly) 'true justified belief', the notion of a 'new type of knowledge' begins to look incoherent. New types of fact/truth crop up either never, or every minute, belief stays pretty much the same, and only the justification is sometimes novel.

Insofar as the argument is based on a claim of a quite different category of knowledge, it does not work. It does, however, draw attention to a central feature of consciousness, the so-called 'qualia', but they are as much a puzzle for the dualist as for the reductionist.

3

Nagel's puzzle and challenge of 'What is it like to be....?' (1974) seems to raise two problems for the reductionist: the striking privacy of mental life, and the odd fact of subjectivity (that something should have a 'point of view').

The privacy question was elegantly captured by Leibniz in his vision of the brain expanded to the size of a flour mill, where thought was nowhere to be seen (1714: § 16). However, if we examined the granular structure of an iron bar, would we see the magnetism? Or learn from a highly magnified view of a bird's feather that it was flying? Reductive physicalists would say that Leibniz was examining the brain at the wrong level if he wanted to see thought.

To see the subjectivity of bats as a problem requires imagination, but that is a dangerous tool in philosophy. It is tempting to think that chimpanzees, rats and slugs have either got what we have got, or they haven't. But Dennett makes a good case for saying that animals may be much less conscious than we imagine (1996), so that Nagel may have slipped into the anthropomorphic view he was keen to avoid. Even small insects have eyes, and make decisions based on their own interests, but they have limited nervous systems, and may not be conscious at all.

Subjectivity may well come in degrees. If experience can tail off down the evolutionary tree, then we should recognise that full subjectivity is built up from humble ingredients. Subjectivity is a striking phenomenon, but not necessarily a separate ontological category.

4

Central to the claims of Chalmers is his view (1996:96) that zombies are logically possible (that is, that I could have a perfect twin who was identical to me in every physical respect, and yet lacked consciousness). The physicalist slogan is 'the physical facts entail all the facts' (Kim 1996:232), but this would be false if a zombie was possible.

Chalmers's claim seems akin (using an analogy of Kripke's, 1972:153) to saying that God could create the physical world on Thursday, and then add the consciousness on Friday. If God were to add a new substance on Friday, this, of course, would imply traditional dualism; if, however, he were to add a new property on Friday to his completed physical world, this is akin to creating the rocks on Thursday, and then adding their hardness on Friday, but if the rocks

weren't hard on Thursday, then they lacked the causal powers of rocks (and my zombie twin has the same causal powers as me). The remaining option seems to be that God is adding an epiphenomenon, but this too is rejected by Chalmers.

Chalmers is not really offering an argument here, but is making a claim, which is based on the assumption that reductive physicalism is false. The clearest response is to say that the existence of zombies does not look remotely plausible unless traditional dualism is correct. Chalmers rejects dualism, so we need not discuss why he is right about that.

5

Davidson's claim (1970) is that while mind is caused by brain events, there are no possible laws which could account for how our reasons generate our actions, which leaves mental events as 'anomalies' in the physical world. Reductionism is said to require 'bridge laws', and these do not exist.

The word 'supervenience' is used to indicate a relationship which is neither an identity, nor a connection, nor a normal relationship of property and object, but Searle rightly suggests that 'supervenience' becomes a rather empty term if Davidson concedes that mental events are caused by brain events (1992:126).

The crux of Davidson's claim is that reasons cannot be reduced to physics, but their reduction doesn't seem an obvious contradiction. A plant has a reason to put down roots. We may call this an entirely 'external' reason. But a frog has a reason to dart its tongue out at a fly, and this seems to have an 'internal' component in the frog's perceptions and drives. And a chess-playing computer has a reason to move a pawn. It seems reasonable to escalate this account to human reasons, though we must allow for the striking expansion in scope of imagination made possible by language, imagery, mental models, and meta-thought, which must also be included in the reductive account.

Therefore we may reasonably claim (with Fodor 1989:151) that psycho-physical bridge laws are logically and naturally possible. The prospect of discovering them looks daunting, but it is more to the point that such laws are unlikely to be worth the trouble. There is no meta-law (not even Ockham's) that says all laws must be simple. The law that says that the Grand Canyon is a good reason for visiting Colorado is no more worth working out than is calculating where every stone will land after a volcano explodes.

6

It is commonly suggested nowadays that consciousness is a property of the brain (e.g. Searle 1992: 14), so the anti-physicalist consensus has converged on Davidson's implied 'property dualism', the notion that there is one substance, but that its two properties are so radically different as to constitute two disparate realities.

The best account of properties (e.g. Heil 1996: 181) says that they are features of objects which enable them to be picked out, or which give them causal powers. It seems that some properties are naturally necessary, and others are contingent. Thus it seems that if a bar of iron is organised the right way, it must exhibit magnetism. However, we could easily see how a London bus could fail to have the property of being red. In addition, we can distinguish between properties which are independent and those which are mutually entailed (by natural

laws). Thus being green is independent of being square, but being square entails have fairly sharp corners.

However, when Heil comes to define property dualism (1996:48) he can say no more than that they are ‘two distinct families of properties’. The obvious question from the sceptic is to ask for a distinction between ‘property dualism’, and merely ‘having two properties’. Colour and shape seem to be two families of properties, but clearly being red-and-square is not a case of property dualism, and neither is being square-and-sharp, although shape and causal power sound like separate ‘families’.

This seems to leave property dualism in no-man’s-land. The relationship seems to be nomologically necessary in the upward direction, making the mind a physical property, but necessitated by a different type of cause in the downward direction, which suggests a non-physical property. Davidson says the reasons are the causes, but how do they do this in a physical world? We are either left with reasons coming from nowhere, which implies traditional free will, or we must expect a reductionist account of reasons to emerge, as suggested above.

The situation for property dualism will look particularly bad if the ‘neural correlate of consciousness’ is discovered, because if this really is fully correlated with awareness, we would then have a natural law connecting mind and body (either through a ‘Humean’ regularity, or through a proposed natural necessity), so they would become mutually entailed.

7

It seems to me that the foregoing brief discussions make the only tenable account of the mind a thoroughly reductive physicalist one, even if one didn’t already embrace Lewis’s remark that ‘My reductionism about mind began as part of an a priori reductionism about everything’ (1994:291). This promptly points to the question of free will, which hovers in the background of Davidson’s Kantian defence of reason, and is also central to the problem of morality in a physical world.

The traditional arguments for free will tend to be subjective, so we may start with a re-examination of experience. ‘A thought comes when it wants, not when I want,’ wrote Nietzsche (1886:§17). This strikes me as partially correct, as I cannot usually predict my own decisions, and though I may follow my own resolve to keep thinking about a topic, I then find myself a passive victim of what then occurs to me about it. Searle proposes (2001 :65) that we experience an undeniable, subjective and free ‘gap’ which falls between the final survey of our reasons for an action, and the actual initiation of the action, but this will not get round Nietzsche’s ‘surprise’ argument, as even Searle will have no idea why he finally ate one brussels sprout rather than another.

For Descartes the striking feature of the will was its feeling of seemingly unbounded ‘space’ (1641:136). But a feeling of freedom only requires that the scope of thought be very extensive, not transcendently infinite. Animals in a very large game reserve have all the freedom they could ever want. And the exceptional scope provided by consciousness, somehow generated within some 10¹⁴ connections in our frontal lobes, gives us huge imaginative freedom. Getting something stupendously large but finite from a few building blocks and rules is not magical, since our language illustrates how to do it. We have a capacity for meta-thought and meta-meta-thought, but this capacity also runs out after a while.

The traditional concept of free will is of something utterly outside all causal laws. An act of will is an event which originates in itself, and has no prior cause. The only non-human analogy we have is the concept of God as the First Cause. Thus we find humans literally seeing themselves as being like God, but this is a huge metaphysical claim, and it doesn't seem persuasive that a mundane mammal contributes a First Cause to reality every time it decides to pick up a cup.

It seems reasonable to conclude that the concept of total 'freedom of the will' is an optimistic delusion, probably rooted in a failure to fully recognise that we are just animals, and that our mental life is reducible to our neuronal life.

8

Is the rejection of free will the moral catastrophe dreaded by the 'neo-dualists'? The most illuminating discussion for our purposes is by David Hume (1748:§VIII), who rejects free will (calling it a 'false sensation or seeming experience' (p.94)), but nevertheless offers a defence of 'responsibility'.

Traditionally we have two concepts of responsibility: causal and personal. Lightning is 'causally responsible' for damaging a tree, but if a vandal does the same thing they are 'morally responsible', which permits blame. In the case of lightning we are willing to track into the past the chain of causation, but we are reluctant to do this with the vandal. We seem to need the metaphysical 'uncaused cause' in order to justify our blame. To look for the causes of the vandal's behaviour seems to be a search for excuses for wickedness, and to undermine our moral commitments.

Hume's solution is to collapse the two types of 'responsible' into one. We simply say that a person is 'responsible' for an action if they caused it, and we only treat such actions differently because they are caused by a very distinctive thing - a human person. Hume must then distinguish between actions for which people are held responsible, such as theft, and actions for which we can find an excuse, such as kleptomania. His solution is to attribute normal human responsibility to actions which are fully caused by the 'character and disposition' of a person (p.98), which is presumably the centre of a healthy and mature consciousness. Actions which can be excused arise from illness, or immaturity, or from peripheral aspects of consciousness, such as sleep-walking. We can go on to deal in blame and punishment or their opposites simply because the cause of the action is a very distinctive thing (a person), the behaviour of which can be influenced by reasoning and training (unlike lightning).

However, Hume's whole picture depends on a person having a 'character'. This has to be equated with the 'self', but Hume is famous for his heroic efforts to formulate a strictly empiricist account of personal identity, which started as flat scepticism, progressed to an associationist theory based on 'associations' between ideas to form a coherent whole, and ended in failure, with the admission that no account of a stable core to consciousness could be generated by the theory (1739/40:I.IV.6 & Appendix).

Hume gives us a good account of personal responsibility without free will, but it needs a Self. He is left with a 'natural belief' in some sort of self or character, but with no theory to give the belief stability. A drift away from a true commitment to the Self seems inevitable, starting with

his label of the self as a mere "bundle", and probably moving towards Dennett's 'instrumental' account (1984).

9

Modern brain research, rather to the surprise of traditional empiricists, is homing in on exactly the sort of central controlling self that is required to fill the gap in Hume's theory. For example, Gerald Edelman (2000) says that 'higher order consciousness is accompanied by a sense of self' (2000: 102), which he associates with 'a set of reentrant connections established between language centres and conceptual centres' (p.110). Rita Carter suggests (1998:22) that children develop an 'internal executor' as they become self-conscious. Antonio Damasio talks of the development from 'the simple core self to the elaborate autobiographical self' (2000:134). Precise localities in the frontal lobes are being identified as the areas which produce this phenomenon.

We can (as Russell suggested; 1912:28) concede the existence of such a self at a given instant, without entering into the question of stable identity over time. The neuroscientists do not offer us enough to justify a belief in immortality, or in moral identity over a long lifetime, but it does seem to be enough to fill the hole in Hume's account.

The account of mind, consciousness, freedom and the self which seems to emerge from this sequence of thought is sufficient to allay the fears of 'neo-dualists'. I would suggest that a belief in a physicallybased consciousness, in which a core Self generates characteristic actions, and in which very wideranging thinking can occur, is a sufficient mental basis on which to build a moral theory based on the concept of virtues of character and disposition. Morality is concerned with the actions caused by the Self at the heart of a fully conscious, mature and healthy person, and virtue and vice are judged according to the potential and capacities of that person.

Bibliography

- Carter, R. 1998. *Mapping the Mind*. Phoenix Books.
- Chalmers, D.J. 1996. *The Conscious Mind*. OUP.
- Damasio, A. 2000. *The Feeling of What Happens*. Vintage Books.
- Davidson, D. 1970. 'Mental Events'. In *Essays on Actions and Events*, D.Davidson, 207-227. OUP, 1982.
- Dennett, D.C. 1984. *Elbow Room: the varieties of free will worth wanting*. MIT Press.
- Dennett, D.C. 1996. *Kinds of Minds*. Phoenix Press.
- Descartes, R. 1641. *Meditations on the First Philosophy*, ed. F.E. Sutcliffe. Penguin, 1968.
- Edelman, G.E. and Tononi, G. 2000. *Consciousness. How Matter Becomes Imagination*. Penguin.
- Fodor, J.A. 1989. 'Making Mind Matter More'. In *The Philosophy of Mind*, ed. B.Beakley and P.Ludlow. MIT Press, 1992.
- Heil,J. 1998. *Philosophy of Mind*. Routledge.
- Hume, D. 1739/40. *A Treatise Concerning Human Understanding*, ed E.C. Mossner. Penguin, 1969.
- Hume, D. 1748. *An Enquiry Concerning Human Understanding*, ed L.A. Selby-Bigge and P.H. Nidditch. OUP, 1975.
- Jackson, F. 1986. 'What Mary Didn't Know'. In *The Nature of Mind*, ed. D.M. Rosenthal. OUP, 1991.

- Kim, J. 1996. *Philosophy of Mind*. Westview Press, 1998.
- Kripke, S.A. 1972. *Naming and Necessity*. Blackwells, 1980.
- Leibniz, G. 1714. 'Monadology'. In *Philosophical Writings*, ed. G.H.R. Parkinson. Dent, 1973
- Lewis, D. 1994. 'Reduction of Mind'. In *Papers in Metaphysics and Epistemology*, D.Lewis. CUP 1 999.
- Nagel, T. 1974. 'What is it like to be a bat?'. In *Mortal Questions*, T.Nagel, 165-180. CUP, 1979.
- Nietzsche,F. 1886. *Beyond Good and Evil*,tr.R.J.Hollingdale. Penguin,1973.
- Perry,J. 2001. *Knowledge, Possibility and Consciousness*. The MIT Press.
- Russell, B. 1912. *The Problems of Philosophy*. OUP, 1967.
- Searle, J.R. 1992. *The Rediscovery of the Mind*. MIT Press.
- Searle, J.R. 2001. *Rationality in Action*. MIT Press.
- Smart, J.J.C. 1959. 'Sensations and Brain Processes'. In *Materialism and the Mind-Body Problem*, ed. D.M.Rosenthal, 53-66. Prentice-Hall, 1971.