Moral Functionalis, Supervenience and Reductionism

Frank Jackson; Philip Pettit

MORAL FUNCTIONALISM, SUPERVENIENCE AND REDUCTIONISM

BY FRANK JACKSON AND PHILIP PETTIT

In our paper 'Moral Functionalism and Moral Motivation'[1] we assume cognitivism, the view that ethical language makes claims about how things are, and seek to make plausible under this assumption a view of moral language modelled on David Lewis' treatment of theoretical terms. As one would expect, this gives a theory with many affinities to analytical functionalism about the mind, and we describe our view as moral functionalism. We argue that moral functionalism gives a plausible account of the content of moral statements and beliefs, is supported by the distinctive way that the ethical supervenes on the descriptive, and allows an appropriately intimate connection between moral belief and motivation. Mark van Roojen's discussion (above, pp. 77–81) of MF finds the idea of treating ethical terms as theoretical terms *à la* Lewis attractive, but doubts that we 'have succeeded in offering a reduction of evaluative properties to natural properties' (p. 77); he claims in particular that our moral functionalism does not entail such a reduction. His worries turn partly on problems to do with specifying the natural properties, and partly on a distinctive feature of Lewis-style treatments of theoretical terms. We think that he raises interesting questions but does not touch anything we say in our paper.

Reductive naturalism

Van Roojen thinks of reductionism in ethics as the view that evaluative properties are natural properties. We shall follow him in this usage. Others are possible, of course, and in fact we did not use the term in MF partly because of the misunderstandings that experience has taught us attend its use. Reductionists in this property-identity sense need to tell us what they have in mind by natural properties. They cannot mean properties distinct from evaluative properties, for that definition would contradict the conclusion they are after. They cannot mean properties of things and happenings in the natural world, for that would make G.E. Moore (!) a reductive naturalist. He thought that goodness was a property of happenings in the natural world. Van Roojen wonders if they (we) might mean roughly what is meant in the philosophy of mind by natural properties, but notes that on this proposal the ethical does not *a priori* supervene on the natural. We might have identity in all respects that are counted as naturalistic in the philosophy of mind but still have a moral difference, on account of differences in the kinds of properties dualists believe in – enough agony realized in Cartesian stuff will change moral nature.

[1] *The Philosophical Quarterly*, 45 (1995), pp. 20–39 [hereafter MF].

It is problems like these which in part explain why we never use the terms 'natural property' or 'naturalism' in MF. We talk of *descriptive properties*, meaning properties ascribed by descriptive language, not of natural properties. (The misunderstandings that arose from Jackson's use elsewhere of this term,[2] meaning by it what we here mean by 'descriptive property', played no small part in our avoiding it in MF.) Thus, for us, reductive naturalism (in van Roojen's property-identity sense) is the doctrine that evaluative properties are one and the same as the properties ascribed by certain descriptive predicates, i.e., that the set of properties ascribed by evaluative language is a proper subset of the set ascribed by descriptive language. (And this doctrine we do defend; but more on that in a moment.)

Of course, how to characterize descriptive language is itself an issue in the debate. Perhaps the best approach is to start by culling the ethical and evaluative terms and then specify the descriptive terms as those that are left. This approach makes transparent the disagreement with Moore. For it is clear that he denied that the properties ascribed by ethical terms like 'right' and 'bad' are one and the same as those picked out by some combination or other of non-ethical terms. (Note that we say 'ascribed', not 'denoted'. It is common ground that descriptive terms can denote ethical properties. For example, 'the property Moore most discusses in his first book' denotes the property of being good – that is, denotes the property ascribed by 'is good'.)

We grant that there is much more to be said on the division between ethical and descriptive language, but for our purposes in MF we rested with the point that there is considerable agreement about which terms count as descriptive and which count as ethical and evaluative, and noted that we had to start somewhere. Our paper was directed to cognitivists who accepted that there was some worthwhile distinction to be drawn at the level of language, and to non-cognitivists to the extent that we hoped that moral functionalism would be found sufficiently appealing in its own right to cause some of them to change sides.

With this clarification behind us, we can consider van Roojen's objection that the nature of Lewis-style treatments of theoretical terms means that moral functionalism does not entail reductive naturalism. We start with a brief statement of the essential idea behind moral functionalism. We keep this short as we are going over material already in the papers under discussion.

Moral functionalism

Moral functionalism holds that the meanings of ethical and evaluative terms are given by their place in a complex network. We call this network 'mature folk morality'. It is the really central, albeit partly implicit, part of the rich story that interconnects descriptive and evaluative ways of describing matters; the part that will or would survive critical reflection; the part that we folk consider defines our subject. If we think of the network as containing the ethical terms written in

[2] F. Jackson, Critical Notice of Susan Hurley's *Natural Reasons*, *Australasian Journal of Philosophy*, 70 (1992), pp. 475–87. For G.E. Moore's worries about what 'non-natural' means, in his claim that goodness is a non-natural property, see 'A Reply to My Critics', in P.A. Schilpp (ed.), *The Philosophy of G.E. Moore* (Northwestern UP, 1942), esp. pp. 581–92.

property-name style, as, for instance, saying that an action's having rightness implies its having a greater share of goodness than at least one alternative action, we can think of the network as saying that it is *a priori* that rightness is the property that plays the role specified – and likewise for all the evaluative properties. More precisely, this is the right way to think of the account if you take the realizer and not the role property option. As we explain in MF, a moral functionalist might think of rightness not as the property that plays the rightness role but as the property of having a property that plays the rightness role. If M is mature folk morality written out in the property-name way, then M is of the form '... V_1 ... V_2 ... V_n ...', where the V_i are the names of the various ethical properties; they are the evaluative terms. We can then give the conditions under which an act is right, that is, has rightness, thus:

R.   *a* is right iff $(\exists x_1)(\exists x_2)$ ... $[M(x_1, x_2, ...) \& a$ has $x_r]$

where $M(x_1, x_2, ...)$ is M with the evaluative terms replaced by free variables, and 'x_r' replaces 'rightness'. (We simplify by ignoring the issue as to whether it is part of folk morality that there is only one, possibly disjunctive, property occupying the role definitive of each ethical property.) As the evaluative terms are all replaced by variables, the right hand side of (R) is free of value expressions. Moreover, according to moral functionalism, the truth condition for '*a* is right' as given by (R) is simply the formal expression of the idea that rightness is defined by its role in the network M; thus, according to moral functionalism, (R) is *a priori*. In the same way there will be *a priori* biconditionals for '*a* is good', '*a* is evil', etc., whose right hand sides contain no value expressions.

Moral functionalism and reductive naturalism

It follows that moral functionalism is committed to *a priori* true biconditionals linking each sentence of the form '*a* is V_i' with a sentence framed in purely descriptive terms. Thus to the extent that van Roojen is sympathetic to a Lewis-style treatment of the evaluative terms, he should be sympathetic to a view that represents a sort of reductionism. However, it is quite another question whether the properties whose places in the network can be given in purely descriptive terms are *themselves* descriptive properties. For all that moral functionalism says, it might be that the properties whose descriptively specifiable interconnections make it the case, according to moral functionalism, that they are the evaluative properties are not themselves descriptive properties – just as van Roojen says.

Nevertheless, we did suppose that the properties that are so interconnected are themselves descriptive properties; we did subscribe to reductive naturalism in that sense. We did so, however, not because of our adherence to moral functionalism but because of our adherence to the distinctive nature of the supervenience of the ethical on the descriptive.

We took it (from our cognitivist perspective) that the ethical supervenes on the descriptive in the following sense

S.   For all *w* and *w'*, if *w* and *w'* are exactly alike descriptively, they are exactly alike ethically.

(S) is a global supervenience thesis, one that quantifies over complete ways things might be. (Of course, non-cognitivists will make little sense of (S), since on their view talk of worlds being alike ethically makes no literal sense.) This is not the kind of intra-world supervenience thesis often discussed in connection with universalizability in ethics. (S) is of course compatible with the view that ethical nature, the ethical way things are, is in part determined by facts about our (and other sentient creatures') responses and attitudes. For included in the global descriptive supervenience base will be facts about responses, both actual and hypothetical, and both first- and higher-order.

If (S) is true, any ethical sentence, in the sense of any sentence about ethical nature, is logically equivalent to some purely descriptive sentence. Let E be a sentence about ethical nature in the following sense: (a) E is framed in ethical terms and descriptive terms; (b) every world at which E is true has some ethical nature; and (c) for all w and w', if E is true at w and false at w', then w and w' differ ethically. The idea is that E counts as being about ethical nature by virtue of the fact that there must be some ethical nature for it to be true, together with the fact that the only way to change its truth-value is by changing ethical nature: the worlds must, that is, differ somehow in the distribution of ethical properties and relations. Now each world at which E is true will have some descriptive nature: ethical nature without descriptive nature is impossible. And for each such world there will be a sentence containing only descriptive terms that gives that nature in full. Now let w_1, w_2, etc., be the worlds where E is true, and let D_1, D_2, etc., be purely descriptive sentences true at w_1, w_2, etc., respectively, which give the full descriptive nature of w_1, w_2, etc. Then the disjunction of D_1, D_2, etc., will also be a purely descriptive sentence: call it D. But then E entails and is entailed by D. For every world where E is true is a world where one or other of the D_i is true, so E entails D. Moreover, every world where one or other of the D_i is true is a world where E is true, as otherwise we should have a violation of (S): we should have worlds which are descriptively exactly alike differing in ethical nature. Therefore D entails E.

What (S) means, therefore, is that any claim about how things are ethically is equivalent to some claim, frameable in purely descriptive terms, about how things are. This is not a *proof* that ethical properties are one and all descriptive properties. But it does, we submit, establish a very strong case that they are. Despite the notoriously *a posteriori* nature of 'x is water iff x is H_2O', few think that water and H_2O are distinct kinds. True, some think that although 'x is an equilateral triangle iff x is an equiangular triangle' is necessarily true, the property of being an equiangular triangle is distinct from that of being an equilateral one. But on examination their arguments bear not so much on properties, in our sense of ways things might be, but on a notion more closely tied to the *concepts* involved.

Van Roojen sees it as part of our project to defend reductive naturalism via moral functionalism. However, we never thought, and do not say, that reductive naturalism follows from moral functionalism. Rather, the identity of ethical properties with descriptive ones is strongly supported (given cognitivism) by the distinctive nature of the supervenience of the ethical and evaluative on the descriptive. The challenge that arises then is to give an account of how it might be that ethical

terms pick out purely descriptive properties. Moral functionalism is an answer to this subsequent question.[3]

Australian National University

[3] We are indebted to discussions with David Lewis and Michael Smith, and to a query by Mark van Roojen about an earlier draft.

CONDITIONALS, FUNCTIONAL ESSENCES AND MARTIN ON DISPOSITIONS

BY STEPHEN MUMFORD

I. CONDITIONALS AND THE DISPOSITIONAL/CATEGORICAL DISTINCTION

What is the relationship between disposition ascriptions and conditionals? There has recently been a significant contribution[1] to this old debate that has succeeded in bringing some important issues into sharper focus. C.B. Martin's conclusions that disposition ascriptions are not *equivalent* to stronger-than-material conditionals is a desirable one if we want to take seriously the notion of dispositions as real *properties*. If it was possible that a disposition ascription could be analysed entirely in terms of a (set of) stronger-than-material conditional(s), then dispositions would be reducible to (complexes of) events described in such a (set of) conditional(s) and could drop out of our ontology altogether.[2]

   If Martin's argument succeeded, it would establish realism about dispositions. However, it is disturbing in a number of respects. The chief among these relates to the dispositional/non-dispositional distinction and the implications for any proposed dispositional realism. Should the argument be accepted, there would seem to be no basis on which we could distinguish the dispositional from the other categories from which it is traditionally set apart, such as 'structural', 'categorical' and 'occurrent'. If so, then the prospect of replacing a conditional analysis of disposition ascription with 'first-order dispositions or powers' (Martin p. 7) seems hopeless, for it is unclear what he is asserting to exist. The problem concerns what makes any property *P* a dispositional rather than a non-dispositional or categorical property. (I am in agreement with Martin (in conversation) that the term 'categorical' is seriously misleading as a contrast with 'dispositional'. 'Categorical' just means 'actual', but an ascription of a

[1] C.B. Martin, 'Dispositions and Conditionals', *The Philosophical Quarterly*, 44 (1994), pp. 1–8.
[2] This reductionist line has an origin in empiricism; see for example J. Bricke, 'Hume's Theory of Dispositional Properties', *American Philosophical Quarterly*, 10 (1975), pp. 15–23.