



**WILEY-
BLACKWELL**

Moral Functionalism and Moral Motivation

Author(s): Frank Jackson and Philip Pettit

Source: *The Philosophical Quarterly*, Vol. 45, No. 178 (Jan., 1995), pp. 20-40

Published by: Blackwell Publishing for The Philosophical Quarterly

Stable URL: <http://www.jstor.org/stable/2219846>

Accessed: 13/10/2008 12:55

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=black>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.



The Philosophical Quarterly and Blackwell Publishing are collaborating with JSTOR to digitize, preserve and extend access to *The Philosophical Quarterly*.

<http://www.jstor.org>

MORAL FUNCTIONALISM AND MORAL MOTIVATION

BY FRANK JACKSON AND PHILIP PETTIT

Cognitivism is on solid ground, it seems to us, when they maintain that practical evaluations – ‘this is right’, ‘that is fair’ – are expressions of judgement and belief (even if these may be massively mistaken). Their non-cognitivist opponents deny the appearances when they say that such evaluations are expressive only of feeling or desire or some such non-cognitive attitude that differs from belief and its cognates in not being subject, in its own right, to the same disciplines of evidence and logic and reasoning.

But while cognitivism is on the side of common sense in this respect, they have had a hard time meeting two challenges (see Blackburn 1984, 1993; Gibbard). The first challenge is to say what the cognitive content of evaluative belief is, given that the content is supposed to give direct if defeasible support to a conclusion about what should be desired and done. The second challenge is to explain why assent to such a content is not just inductively associated with forming a corresponding desire to act as the evaluation prescribes, why it is a matter of surprise, to be explained by some *malaise* like weakness of will, if someone sincerely assents to the positive evaluation of an option without feeling any inclination towards it (Smith 1989). The first challenge comes of the Humean assumption that matters of cognition, matters of putative fact, do not imply prescriptions: there is a gap between ‘is’ and ‘ought’. The second comes of the assumption, equally Humean in origin, that states of cognition are motivationally inert and only lead to behaviour under the impulse of a distinct state of desire: reason, as it is said, is the slave of the passions.

This paper is an attempt to sketch an answer to these two challenges. It presents what we describe as a functionalist account of the cognitive content of evaluative belief, and then shows why assent to such a content should generally be expected to involve the presence of a corresponding desire. We stress that the paper is only a sketch of our answer to the two challenges, for the matters we discuss connect with almost every issue in meta-ethics and it is not possible to do full justice to them here. We offer the sketch in the hope that the position we identify, once it has

been clearly discerned, will look as simple and attractive to others as it looks to us.

There are two main sections to the paper, corresponding to the two challenges that we have distinguished.¹ The first gives an account of the cognitive content of evaluative belief. The second explains why the belief in such an evaluative content should generally involve a corresponding desire. We assume that the category of evaluative content is well enough identified by examples such as 'that is right', 'this is fair', and so on, but we say a little more about the category at the end of the first section.

I. THE CONTENT OF EVALUATIVE BELIEF

In this part our aim is to present the functionalist theory of evaluative content. The argument for the theory has two premises, one bearing on the supervenience of the evaluative on the descriptive, the other on the networked nature of moral terms. We shall present each of the premises, characterize the theory that they support, and then show why the sort of content in question counts as evaluative.

The supervenience of the evaluative on the descriptive

One of the most striking features of moral beliefs, and evaluative beliefs generally, is that the evaluative facts they bear upon satisfy a certain supervenience condition relative to descriptive facts. It is usually urged that if two options are descriptively indiscernible in regard to universal features then they are morally indiscernible as well: supervenience is associated with universalizability. This supervenience of the evaluative on the descriptive is *a priori* knowable, because it is marked by a common feature of our practice in moral argument, and anyone party to that practice is in a position to recognize that the supervenience obtains. The feature is that you can justifiably make a moral difference between two options only so far as you can point to a descriptive difference between them. No evaluative discriminability without descriptive discriminability. If you say that that action is right, but this wrong, and you cannot point to some other, descriptive difference between them, then your status as a participant in evaluative discourse will be up for question.

The supervenience of the evaluative on the descriptive, as we have

¹ The main elements of the argument in §I are in Jackson 1992, those of §II in Pettit 1991 and 1993.

characterized it so far, means that for any possible world w , if x and y are descriptively exactly alike in w , then x and y are alike in moral respects too. But our practice in moral argument supports a somewhat stronger supervenience claim, and makes it *a priori* knowable too: for any two worlds that are descriptively exactly alike in other respects, in the attitudes that people have, in the effects to which actions lead, and so on, if x in one is descriptively exactly like y in the other, then x and y are exactly alike in moral respects. More globally, if two worlds, w and w' , are exactly alike in descriptive respects, then w and w' are exactly alike in evaluative respects too. The fact that moral evaluation is answerable to descriptive considerations means that there is no room for a difference of value between two worlds that are exactly alike in descriptive regards.

The *a priori* supervenience of the moral on the descriptive gives sharp expression to the lesson taught by the answerability of evaluation to descriptive considerations. This is that there is no possibility that moral beliefs might partition descriptively indiscernible possible worlds in a manner additional to the partitioning introduced by descriptive beliefs. If two worlds are descriptively indiscernible, then there is no difference available to be captured in evaluative terms. Characterize a world or an option evaluatively and you assign it to a sort that is adequately identifiable in descriptive terms. Evaluation, to put an ironic twist on the lesson, is description by other means.

The networked character of moral terms

What, then, are we to say about the meaning of moral terms? The descriptive answerability of evaluation would be unsurprising, if evaluative terms were all descriptively definable, one by one. But we see no plausibility in the idea of providing such an atomistic reduction of evaluative terms. Nor is the idea likely to appeal to many contemporary philosophers. By most accounts, moral terms are involved in a network of content-relevant connections, including connections with other moral and evaluative terms (see Hurley; cf. Quine 1953, 1974). No simple atomistic definition is going to yield an understanding of a moral term, because each such term is used in a way that presupposes a large network of connections with other terms, evaluative and descriptive. Take the example of fairness. That something is fair means that conditions of the kind registered in the following illustrative commonplaces are fulfilled.

1. *Commonplaces about presentation.* In most contexts 'I cut, you

- choose' is a fair sort of procedure: if anything is fair, then that is. And a similar exemplary status is enjoyed by equal division and by any procedure that departs from equal division only so far as independent claims require. Other arrangements are judged to be fair on the basis of similarity with such paradigms.
2. *Commonplaces about truth-conditions.* There can be no difference in fairness without a descriptive difference: fairness is descriptively supervenient.
 3. *Commonplaces about justification.* If one alternative is fair, and if other things are equal, then that is the right option for the agent to desire and pursue. The agent may reasonably feel guilt, and others resentment, about failure to choose such an option, at any rate where the choice would not have been very difficult to make.
 4. *Commonplaces about justificatory power.* Fairness is potentially more important in the determination of rightness, and in the justification of choice, than being polite or diverting. But fairness is less important in general than saving innocent human lives: better be unfair than allow someone innocent to perish.
 5. *Commonplaces about motivation.* Anyone who believes that one option is fair will at least prefer that option, *ceteris paribus*, to alternatives. Not to do so is probably to be put down to a failure like weakness of will, a form of practical unreason (Pettit and Smith).
 6. *Commonplaces about motivational power.* Believing one option to be fair is likely to motivate an agent more strongly than seeing another as polite or diverting, but less strongly than recognizing a further option as a means of saving innocent human life. If these asymmetries do not hold, that is probably due to some form of practical unreason.
 7. *Commonplaces about virtue.* The fact that something is fair, if it is a fact, is likely to be more salient to people the more they display the virtuous character traits of spontaneously choosing what is right.

Such commonplaces are candidates for *a priori* truths: they are putatively such that anyone who knows how to use the term 'fair' is in a position to see that they hold. We may expect debates, of course, about which commonplaces are indeed *a priori* true, and we return to this issue later. Those that are *a priori* true will give us the conditions under which fairness is instantiated, though it should be noticed that they do not entail that the property ever is instantiated. As 'Bachelors are unmarried

males' is consistent with there being no bachelors, so the commonplaces for fairness are consistent with there being no such property realized in the actual world.

What is true of 'fair' is true of any moral term. The term makes its particular contribution to evaluation just so far as it engages the commonplaces linking it with other terms. What is said or believed in characterizing an option as fair ties up with the fact that accepting the characterization means becoming disposed to draw the lessons articulated in the commonplaces. If I think that an option is fair then I expect people, myself included, to be able to see it as of a kind with certain paradigms of fairness; to be more impressed by that fact than by the fact that it is a somewhat embarrassing choice to make; to believe that, other things being equal, the option is right; and, if I am not the victim of some *malaise*, to be drawn towards the option. It is in virtue of making such connections, at least in part, that the characterization of the option as fair is contentful or informative.

The networked character of moral terms means that there is no plausibility in the idea of giving each term a separate descriptive definition, without regard to the definition of others. What it is to be fair, in descriptive terms, must tie up with what it is to be right, in descriptive terms, if a connection between those properties is assumed in the way we deploy the words. But if moral terms are not going to be definable, one by one, in descriptive terms, how are we to make sense of the answerability of evaluation to descriptive considerations? How are we to make sense of the *a priori* supervenience of the evaluative on the descriptive?

The functionalist theory of moral content

The most attractive possibility that we see is to view moral terms as theoretical terms in the style of David Lewis (1983 Essay 6; see Jackson 1992, 1994). The moral terms are specified by their role in received moral theory – in 'folk moral theory' – and while this theory has a purely descriptive content, so that we can understand the answerability of evaluation to descriptive considerations, the content of any one claim is fixed only so far as the contents of others are fixed simultaneously. Moral terms are reducible to descriptive terms, at least in principle, but the reduction involved is holistic, not atomistic.

This is the idea (see Jackson 1992 for a slightly more formal presentation). What property do I ascribe to an option in characterizing it as fair? I ascribe to it that property which fills the place marked out for fairness in ordinary moral thinking: it belongs to certain paradigms that we find saliently similar; it inclines us to judge that a bearer is right,

i.e., has the descriptive property associated with the rightness-role; it inclines us to make this judgement more or less strongly than would certain other descriptive properties that are assigned otherwise similar roles within the theory; it tends to arouse a desire for the realization of the option; and so on. The term 'fair' picks out a descriptive property, then, but it does so by virtue of the place that that property occupies in folk moral theory, and in a manner that requires other moral terms simultaneously to pick out complementary descriptive properties.

The approach characterized in these remarks amounts to what we call a functionalist theory of moral terms. The analogy with the familiar functionalist theory of psychological terms will be obvious. According to the psychological functionalist, to believe that *p* is to instantiate an as-it-happens physical state of displaying a belief-that-*p* profile: it is to exhibit the required connections. According to the moral functionalist, for an option to be fair is for it to instantiate a descriptive property such that the option displays fairness-behaviour: it connects up with other things in the manner distinctive of fair alternatives.

But while the psychological parallel may be useful in introducing moral functionalism, we should stress the differences between the doctrines. First, it is neither necessary nor *a priori* knowable that physical properties are what ensure the fulfilment of the roles associated with psychological properties. It is, however, both necessary and *a priori* knowable that descriptive properties are what ensure the fulfilment of the roles associated with moral properties. Second, functionalism about the mind associates primarily causal roles with the various mental properties. By contrast, although the functionalist account of moral properties may include as part of the role definitive of fairness 'Many acts are right because they are fair', the 'because' is not a causal 'because'. The fairness of an action does not causally explain its rightness.

We think that a functionalist account of moral content is attractive for its ability to explain the supervenience of the evaluative on the descriptive, and the *a priori* status of that supervenience; in particular, it can explain these consistently with admitting the networked character of moral terms. If two worlds are descriptively indiscernible, then they will be indiscernible in regard to properties associated with roles that descriptive properties underpin: same underpinnings, same roles. And so the truth of moral functionalism would ensure the (*a priori*) supervenience of the evaluative on the descriptive.

There is more to be said, however, in favour of the functionalist view. It makes good sense of what goes on in moral thinking, both in everyday life and in philosophy. In moral thinking we try to form our opinions on new moral questions and to examine and, if necessary, revise our

existing opinions on old. According to the functionalist view, the meaning of relevant moral terms will be fixed by roles which certain commonplaces give them, and so moral thinking is bound to involve the attempt to use commonplaces as a base, and holding on to as much of that base as possible, or at least to the parts considered most secure, to fix opinions on particular questions. The commonplaces that emerge in this process as those whose rejection cannot be countenanced will be taken to fix the relevant roles; they are the *a priori* compulsory propositions that anyone who knows how to use the terms is in a position to recognize as true. Other commonplaces – other putatively *a priori* propositions – will have to be dismissed as false or downgraded to the status of empirical, contingent truths.

This account of moral thinking fits with Rawls' influential argument that systematic moral thinking involves the attempt to equilibrate such general principles as we find we really cannot give up with considered judgements as to how various options should be morally characterized. Is such and such an arrangement to be regarded as fair? The issue turns, by his account, on whether judging that it is fair is consistent with general considerations like the following: that its fairness ought to put it in a salient family resemblance with certain paradigms (is it really on a par with equal division?); that its fairness ought to mean that we would desire it if unbiased (would parties in the original position prefer it?); that its fairness ought to make it desirable in a more powerful and important way than would the property of being, say, economically efficient (does it have the 'primary' virtue of an institution?); that its fairness ought to make it look like the right arrangement to have (can we give it our considered approval?); and so on. This is exactly what we would expect, under a functionalist account of moral content.

The fact that moral thinking follows this pattern means, it should be noticed, that folk moral theory is always on the advance, as people look for the most consistent way of equilibrating their general principles and their particular judgements, not just about fair institutions, but also about whether abortion should be available on demand, whether the state should impose any moral views and whether the public water supply should be fluoridated. The drive for equilibrium or consistency may reflect back on general principles, as Rawls himself emphasizes: it may make for a shift in our shared folk theory.

The functionalist account also makes sense of the more comprehensive philosophical enterprise that attempts, not just to develop systematic moral opinion, but to identify a canonical method of moral thought. Different moral methods, as we see them from a functionalist viewpoint, start from different points in the moral network of putatively *a priori*

commonplaces, and argue that their chosen starting-point offers the most advantageous platform for moral thinking, the platform from which it is easiest to advance moral opinion, while preserving as much as possible of the network. Utilitarians start from the commonplace connection of value with certain descriptively given paradigms, in particular with the realization of happiness, arguing that we can save most of the commonplaces of folk moral theory, and develop reliable moral opinions, by looking always for what promotes happiness. In a similar vein, subjectivists and dispositionalists, perhaps non-cognitivists too, start from the commonplace connection between value and what we desire or would desire in favoured circumstances; contractarians start from the commonplace connection between the positive evaluation of something and its being found justifiable by others; virtue theorists start from the commonplace connection between the perception or pursuit of value and the possession of certain character traits; and so on.

This view of existing moral philosophies is highly ecumenical in suggesting that they represent different but not necessarily inconsistent methods or strategies for moral thought. As a matter of fact, of course, most of the philosophies in question are given constitutive as well as methodological significance by their defenders; they are presented as rival, sectarian attempts, in T.M. Scanlon's words (p. 106), to give 'a philosophical explanation of the subject-matter of morality'.

There are two different ways, from a functionalist perspective, to give constitutive significance to a sectarian philosophy. One may hold, consistently with functionalism, that a particular approach points us towards the descriptive properties that fill the moral roles. The role-fillers are all felicitic properties, the utilitarian will say; they all relate to what people find unobjectionable, the contractarian will say; and so on. Alternatively, one may hold that a given sectarian philosophy tells us that only certain parts of the moral roles, only certain subsets of the commonplaces, matter to the strict understanding of the moral terms; the other commonplaces may be accepted or rejected, but in either case they will not be involved in the meaning of moral terms. This alternative will suggest that folk moral theory is excessively generous in the networks of commonplaces that it countenances: all that will allegedly matter for the meaning of this or that moral term is consistency with some fragment of the network – the part that specifies certain paradigms, or a motivational connection, or a justificatory capacity, or a linkage with virtue, or whatever.

In conclusion, there are three things it is important to notice about our functionalist account. First, it leaves open whether rightness, say, is the ground-level (descriptive) property that occupies the rightness role,

e.g., the property of maximizing happiness, or whether it is the higher-order property of having a property that occupies the rightness role (see Jackson and Pettit). On either approach, an action is right if and only if it has that role-filling property: the question left open does not bear on the conditions for being right but on the metaphysical matter of what rightness, the moral property, is. And on either approach, to believe that ϕ -ing is right is not to believe that ϕ -ing has the descriptive, role-filling property, e.g., that it maximizes happiness. It is, rather, to believe that ϕ -ing has the property that occupies the rightness role, whatever that property is.

The second thing to notice is that the version of moral functionalism that identifies the moral properties with the ground-level, role-filling properties comes in two salient varieties. We can think of a role in folk moral theory as picking out a moral property in rigid or in non-rigid fashion. On the first construal, rightness is associated with the property that *actually* occupies the role of rightness; on the second, rightness is associated with a property that may vary from world to world, as now one property, now another, occupies the rightness role *there*. Under either construal, however, it may be *a posteriori* which property rightness is associated with.

The third thing to notice is that the functionalist theory is not exposed to any simple version of the open question argument. Take, for instance, the argument that says it cannot be correct to hold that x is right if and only if x has some role-filling descriptive property, because with any descriptive property it is open to significant question whether or not it is right to pursue that property. This argument has no bite precisely because the biconditional linking rightness with a role-filling descriptive property may be *a posteriori* in character, according to the theory. The theory allows that it need not be a conceptual matter that x is right if and only if x is F , where ' F ' might be 'maximizes happiness', 'accords with such and such a list of *prima facie* duties', 'is what a person with such and such virtues would do', or whatever. What has to be a conceptual matter is only that x is right just in case x has the property, whatever it is, that plays the rightness-role.

It is true that, according to the functionalist theory, the total body of descriptive information entails the evaluative way things are (Jackson 1992). For the network of connections posited in folk moral theory can be given in purely descriptive terms, and the information about which properties fill the roles can also be given in descriptive terms. But the entailment here will not be an obvious one, since the network postulated is complicated and in process of articulation, under the drive for reflective equilibrium; arguably, indeed, it is something we shall never reach

but rather shall converge on over time. Besides, as we shall see in the next section, our folk moral grasp of that network has features that undermine the open question argument. It is tacit in a way that makes claims about the descriptive connections less than immediately accessible. And it is practical in a way that links access to those connections with the presence of a presumptively defensible desire which it is hard to envisage having while at the same time regarding the relevant question of rightness as open.

Evaluative content?

The principal mark of whether a given content has practical evaluative significance, i.e., of whether it is an evaluative as distinct from a purely descriptive content, is whether assenting to the content gives a person reason, however defeasible, to form the appropriate desire and choice: whether the truth of the claim in question offers inferential or quasi-inferential support – deliberative support, as it is usually described – for such a positive response. We think that our account of the content of practical evaluation explains why moral content has this decision-supporting significance.

We can distinguish direct from indirect information bearing on what we think we ought to do. Direct information tells us about, say, the incidence of caries after fluoridation, the costs and the possible risks of fluoridation, and so forth. Indirect information tells us about the attitudes of others to fluoridation, and about what we ourselves would feel if various circumstances were different, or if we knew more, and so forth. The indirect information is not so much information about the nature of the contemplated course of action itself as information relating to attitudes of ourselves and others to the course of action.

Indirect information can be very telling for our views as to what we ought to do. ‘You wouldn’t take that attitude if you hadn’t had that row with Jones last year’, ‘People whose judgement you respect wouldn’t approve’, ‘If you knew what I know, you wouldn’t go ahead’, ‘An expert committee has recommended against it’: there is something amiss about someone who accepts such claims but fails to see them as reasons, however defeasible, for acting accordingly. Suppose you believe that you would desire something if a certain irrelevant incident had not occurred, or if you were as wise as some people you admire, or if you had certain information or expertise. You are then going to believe that if you do not desire it now, that is because of some inappropriate factor. It should be no surprise if in these circumstances you see reason why you should desire the matter in question.

The functionalist account of the content of moral belief shows that it gives the subject the same sort of indirect information about the appeal of the option it recommends. We mentioned, in illustration of the connections that go with fairness, that to judge that an option is fair or right, other things being equal, is to tend to desire it. But this means that to judge that something is fair or right, and that other things are equal, is to have indirect information of the sort that we have been discussing, to believe, e.g., that if ideally situated in the relevant way, one would desire the option. And so one will have reason to think that one ought to desire and pursue the option now.

Is it plausible to posit the connection envisaged in our functionalist account between evaluative belief and desire? We believe so. When I decide that a course of action is the fair or right course of action, I do seem to be committing myself to the view that, other things being equal, this is the sort of action that I would desire to do, or perhaps desire to desire to do, were I not weak-willed, or were I to set aside prejudices and biases, or were I able to think myself into the shoes of those affected, or something of that kind (see Smith 1992). And so it is intelligible that when I believe that a course of action is fair, etc., then I find myself with a reason for desiring it; I contemplate a consideration which tells me that this is what I ought to desire, whether or not I actually desire it.

II. EVALUATIVE BELIEF AND MOTIVATION

So much, then, for the functionalist theory of the content of moral belief. The next problem in the functionalist development of the cognitivist programme of representing evaluations as expressions of belief is to explain why assent to that sort of content is non-inductively associated with the presence of a corresponding desire. We saw in the first part how the content of a belief can inferentially or *quasi*-inferentially support a given course of action and how in this sense it can count as an evaluative content. But a content might be evaluative, might support this or that action, without belief in the content being more than inductively associated with desiring and acting as the content requires. So what explains how it is that believing evaluative contents connects non-inductively with having corresponding desires? In particular, what explains this, given the common assumption that beliefs are motivationally inert?

Responses to this problem in the cognitivist literature vary between two extremes (see McNaughton). At one extreme, some cognitivists say that, when there is no failing of understanding or logic, to evaluate is necessarily to have a corresponding desire, perhaps even a desire of sufficient strength to guarantee appropriate action; this amounts to a

contemporary version of the thesis that knowledge is virtue (McDowell 1978, 1979). At the other extreme, some cognitivists say that evaluation is not particularly closely related to desire: that in this respect, it is on a par with purely descriptive belief (Brink). We shall be defending a middle position. Indeed we think it is a virtue of functionalism that it makes a middle position accessible. We argue that there is an intimate, intelligible connection between evaluation and desire, but that this connection does not mean that people who fail to desire as they evaluate are cognitively deficient.

The problem

On the face of it, functionalism may seem badly placed to make sense of the linkage between moral belief and desire. The account does entail that if I am correct in thinking that something is right or fair or whatever, then I am right in thinking that were I ideally situated I would desire it; and so it entails that if I am correct in thinking this, and if I am indeed ideally situated, then I do actually desire it. But the account has to explain the presence of desire, or the tendency of evaluative belief to be attended by desire, not just under such special conditions, but also when the subject is incorrect or is not ideally situated (Dreier p. 16). And it is not clear how it can hope to do this.

The reason is that under the functionalist account the content of evaluative belief appears to be extremely intellectualized, and belief in such a content looks unlikely to have any particularly intimate association with desire; on the contrary, the very complexity of the content looks likely to distract the most passionate of subjects from desiderative concerns. When I believe that something is fair, I believe on this account that it is similar in various respects to certain paradigms, that the similarity has a certain descriptive basis, that in virtue of this similarity I can justify myself in choosing the prospect in question, that in virtue of this similarity I would desire the option were I ideally situated, and so on. Assent to such sophisticated thoughts is hardly the stuff of passion and action. And so the functionalist account may seem to face a particularly difficult problem when it comes to explaining why evaluative belief is intimately linked to desire.

Why the problem need not be intractable

But if the functionalist account seems to raise a particularly difficult problem, that is because we are assuming that the only way to believe the sort of content associated by functionalism with the claim that some-

thing is fair or right is to contemplate all of the connections in question and then to assent to their obtaining. We need to see that this intellectualist assumption is mistaken, so that the problem is not quite as daunting as it may at first appear. Then we can return to the problem itself and show how it can be resolved.

The main point to make against the intellectualist assumption is that for any of a range of contents there are two broadly different ways in which the content may be believed, one more or less in line with the intellectualist assumption, the other more or less non-intellectual or practical. Three simple examples will suffice to make the point.

Consider the belief that all cats purr. The most familiar way of believing that is to be disposed to assent to the universal proposition, i.e., to maintain the universal belief *in sensu composito*. But another way of holding that belief, at least by traditional accounts, is to have the universal disposition to assent to instances of the original proposition, to be disposed to assent to the claims that Felix purrs and that Boots purrs, so far as you take these to be cats, and so on; this is to maintain the universal belief *in sensu diviso*. Again, consider the belief that it is more probable than not that it will rain. One way of believing this is to endorse the probabilistic proposition given. But another is to believe with a strength of greater than a half that it is going to rain: it is to have a partial belief in the prospect of rain rather than a full belief in the probability of rain. Finally, consider the belief in a principle of reasoning like *modus ponens*. One way of believing this is to endorse the familiar principle according to which it is appropriate to infer that q from premises to the effect that p and that if p then q . But another is to be disposed to draw conclusions in accordance with that principle, without ever spelling out the principle to yourself or representing it to yourself as something that commands assent.

The intellectualist way of holding such beliefs is more sophisticated; it requires concepts that may be lacking in someone who maintains the belief in a non-intellectualist way. That is, it requires the subject to have extra *de dicto* beliefs about what it is for something to hold universally, what it is for a possibility to be probable, and what it is for one truth to entail another. But though the intellectual mode scores over the non-intellectual in this respect, in another way it scores below. For it is possible to believe one of our propositions intellectually but to lack the smooth manifestation of belief in action that characterizes the non-intellectual counterpart. The intellectual believer that all cats purr may not be able to recognize a cat or may only be able to see with difficulty that since Boots is a cat and all cats purr, Boots must purr. The ideal position for the believer in any proposition is to believe it both in intellectual

and in non-intellectual mode; the lack of either mode will involve some deficit.

When are we to say that someone believes a certain intellectualized content in non-intellectual mode? Going by our examples, the answer is: when that person displays dispositions such that it is compulsory, on coming to understand the content, to accept it; in other words, when that person displays dispositions which yield judgements that support the proposition conclusively. Thus the person who on taking something to be a cat is disposed to judge that it purrs is disposed to make judgements that support the proposition that all cats purr. And the person who is disposed to infer the truth of the consequent of a conditional from the joint truth of the conditional and its antecedent is disposed to make judgements that support the principle of *modus ponens*.

The problem raised for the functionalist account of the content of evaluative belief seemed intractable precisely because we were assuming that the only way of believing such a content is the intellectual way. It may be useful to look at one further example of the distinction between believing something in an intellectual and a non-intellectual way. Assume for purposes of illustration that the analysis of colour-beliefs means that to believe that something is red is to believe that the object has a certain more or less objective property that manifests itself in how it looks to the eye, but not invariably: under substandard conditions, or with substandard observers, appearance and reality can come apart. Such a belief-ascription may seem outlandish, since ordinary colour-believers may not pay any explicit attention to observers or conditions, or to their sensations; indeed they may not even have concepts with which to identify such items. But the ascription of the belief becomes plausible once we see it as predicating a non-intellectual mode of belief in the proposition, on a par with the non-intellectual mode of believing a universal or probabilistic proposition. We ascribe the universal or probabilistic belief on the basis of suitable dispositions in the subject, and despite the absence of a concept for the universal quantifier or probability operator. And it turns out that there is a similar basis for ascribing the belief that something has a property of the kind in question, even with subjects who lack concepts for some of the elements in that content.

Consider the following pattern of belief-relevant dispositions that ordinary people might instantiate (Pettit 1993 pp. 92ff.).

1. They identify redness for themselves and for one another in a more or less ostensive fashion, given that bearers of redness have a certain look and that in their presence most participants in

- colour discourse see red. Redness is *that* property, people think, the one present here in this object, there in that, and so on.
2. People typically allow the way things look to determine what else they judge to be red, though they do this without reflecting back on the experience it occasions in them: they do not advert to the sensation of redness.
 3. They are sensitive to discrepancies between their sensations and judgements, baulking at any discrepancy between times or perhaps individuals in whether something looks red and in whether it is taken to be red.
 4. Thus they are ready to discount how something looks as evidence of its being red or non-red; they display a disposition to discount various discrepant appearances in the course of restoring congruence: the light was bad, the person was wearing funny glasses, or whatever.

People with dispositions like this believe, we maintain, that red is a property which tends to show up in how things look, since otherwise they would not go by look; believe that it is an inter-subjective or objective property of things, since otherwise they would not worry about discrepancies of look across times and individuals; and believe that it is that property which shows up in how things look to such subjects and in such conditions as are not apt to be discounted under practices of negotiating about discrepancies, since they would not otherwise put faith in possibilities of negotiation and discounting. We may express this last condition by saying that they believe that anything which is red is such as to look red to normal observers, in normal conditions.

How can it be plausible to ascribe such a complex belief, when the ordinary colour-believer may lack the concept of normality or even the concept of sensation? For the same reason as it may be compelling to say that someone without the concept of the universal quantifier believes a universal proposition. The dispositions of ordinary believers in the redness of something make the complex claim *a priori* compulsory. Given an understanding of that claim, believers are in a position to find, on reflection, conclusive support for it in the judgements that the dispositions incline them to make.

The story about how I may believe in a non-intellectual way that something is red may be described as *ethocentric* (Pettit 1991; 1993 p. 93). It emphasizes unreflective *habits* of response and unreflective *practices* of correction, and the Greek word *ethos* refers to habits or practices; indeed the word 'ethos' has a cognate sense in English. What is crucial to the non-intellectual way of having colour-beliefs is not the theoretical grasp

of any truths but rather more practical factors: the habit of being able to see red, the habit of baulking at discrepancy and the practice of discrepancy-resolution. Post-modernists sometimes use the word *logocentric* in castigating the Western tradition of philosophy: they present the tradition as excessively or exclusively focused on *logos* or theory. The word may be recruited to a sharper usage in describing the intellectual way of holding certain beliefs and in contrasting that mode of belief with the non-intellectual, ethocentric mode that we have been characterizing.

How the problem can be resolved

The problem we raised at the beginning of this section was how an evaluative belief can be non-inductively associated with the presence of desire, given the very intellectual sort of content that it will have under our functionalist account. That problem looked extremely daunting and our discussion of the two different modes of believing a content was designed to make it seem a bit more tractable. The discussion suggests that there is a non-intellectual, ethocentric way of believing that something is fair or right, and that this mode of believing the content may be non-inductively associated with the possession of a corresponding desire. It suggests that non-intellectual evaluative belief may involve desire in the way, for example, in which non-intellectual colour-belief involves sensation, or non-intellectual belief in *modus ponens* involves an inferential disposition.

Take fairness as an example of an evaluative property, as we took redness as an example of a colour-property. And suppose now that people display dispositions like the following.

1. They identify fairness for themselves and one another in a more or less ostensive fashion, given that the exemplars of fair choices have descriptive features that make them attractive for certain purposes and that thereby engage desire: attractive, say, in promising not to generate resentment among those they affect. Fairness is *that* property, they think, the one illustrated by this action, that arrangement, and so on.
2. People find certain other things saliently similar to such paradigms of fairness, no doubt so far as they find them similarly attractive (so the extension of the property of fairness is unlikely to be readily trackable by anyone who lacks the desiderative disposition; see McDowell 1981); and they make judgements of fairness based on such impressions. The similarity registered in

- this way is descriptive similarity, since no difference in fairness is countenanced that does not go with a descriptive difference.
3. People balk at any discrepancy in their impressions about fairness, including any discrepancy in the attraction the property has for them, across times and individuals; they are not prepared to treat such divergence as an insubstantial, no-fault difference (Price p. 161).
 4. They display a disposition to discount one or other of those discrepant appearances: the impression of fairness was misleading, not being based on all relevant information, or being the product of partiality; the non-impression of fairness was due to the absence of a vivid awareness of the option, or to the sort of *malaise* that inhibits the activation of desire; and so on.
 5. When other things are taken as equal, people tend to move from the identification of something as fair to its identification as right or desirable. Alternatively, they tend to move from the identification of the option as fair, and, other things being equal, uniquely attractive, to a view of it as something they can desire without fear of being faulted.
 6. People tend to desire and choose such an option, at least when they explicitly identify it as right or desirable. Anyone who fails to do so will also represent a discrepancy calling for explanation, by reference to contrary impulse, lack of attention to the property that made the option attractive, or whatever.

Under this story about the dispositions of ordinary fairness-believers, certain paradigms get established as exemplars of the property, given the capacity of those paradigms to engage a certain sort of desire. Similarity to those paradigms determines what else is fair, given that it is similarly attractive. But such similarity is not a sure-fire guarantee of fairness: it must be registered in conditions and in subjects fit to survive debate about discrepant responses, for example, under full awareness of what is involved, and in the absence of evaluative *malaise* and bias. It must be registered in what we theorists may want to describe as normal and ideal conditions, by normal and perhaps ideal subjects, 'normal' because certain perturbing factors must be absent, 'ideal' because certain facilitating elements must be present: for example, a vivid awareness of what the alternatives involve. The identification of an option as fair will dispose an agent to see it as right and thus, in the absence of weakness of will, etc., to desire and choose the option.

The pattern of dispositions involved in the story would make it *a priori* compulsory for people who believe that an option is fair to accept that

it satisfies the commonplaces associated with fairness under our analysis, assuming that they come to understand them. The fact that they treat the impression of similarity to paradigms as indicative of fairness, at least when it is motivating, that they allow such an impression of similarity to be definitive only in the absence of discrepancy, and that they tend to move from the identification of an option as fair to the judgement that, other things being equal, it is right: such facts mean that in their mouths 'fairness' refers to the sort of property that satisfies the relevant commonplaces. They may not have the concepts necessary for spelling out those commonplaces, but the way they are disposed to carry on in forming and revising their fairness-beliefs gives them all the evidence needed for supporting the complex claim.

Finally, to the crucial question. Assuming that it is a plausible representation of what is involved in non-intellectually believing that something is fair, does our story explain why such a way of believing that fairness claim is non-inductively associated with having a corresponding desire? We think it does. Under the story presented, people who believe that an option is fair see it as robustly attractive in the same way as relevant paradigms, and do so through treating the desire it occasions, when it is seen in the company of the paradigms, as not liable to be faulted in the course of negotiating discrepancy. They believe that if there were discrepancies between what they desire in this way at different times, or between what they and others desire, then the desires they currently experience would not reveal any departure from the normal or the ideal on their side.

This story means that there is no way of judging non-intellectually that something is fair without experiencing a suitable desire for the option in question. The idea of forming a fairness-belief in this non-intellectual way, and yet lacking the desire, will be as incoherent as the idea of forming a redness-belief in a non-intellectual way and lacking a redness-sensation, or the idea of forming a probability-belief in a non-intellectual way and lacking a suitable degree of confidence. There will be a very intimate connection indeed between forming the evaluative belief and harbouring the corresponding desire.

But intimate though the connection is, we should note that it is not going to be implausibly close. The non-intellectual way of forming the belief that something is fair must count as the non-parasitic or canonical way of making that judgement (for 'canonical', see Lewis 1989). It is hard to see how anyone would make fairness-judgements unless some people made them in this non-intellectual way, as it is hard to see how anyone would make colour-judgements unless some enjoyed colour-sensations and made colour-judgements in a non-intellectual fashion. Our

story allows, however, that one may, without being moved at all, recognize in a purely intellectual way that an option is fair, that it is of a kind with such and such paradigms, and that one would desire it were one not paralysed by some evaluative *malaise* (Stocker). This possibility parallels the possibility of the person whose colour-vision is temporarily or even permanently disturbed but who recognizes intellectually that something is red, perhaps recognizing that others are better equipped or positioned and that they see it as red.

And not only does our story allow that there are intellectual, parasitic ways of forming fairness-beliefs such that no desire may attend the formation. It also lets us see that even when things go quite well, and even when the subject forms a corresponding desire, still they may not go perfectly, and the desire formed may not be as strong as it should be (Pettit and Smith). The subject or the conditions may be good enough to allow the desire to form without satisfying all the constraints associated with the ideal subject and ideal conditions. Bias or self-interest may expose the subject to familiar possibilities of practical unreason.

The nature of moral or evaluative content, as we described it in the first part of this paper, explains why a positive moral evaluation is relevant to whether you should desire something. But it is also consistent with the connection between evaluation and desire being purely inductive. We can now see, however, that where the nature of moral content fails to explain the intimacy of this connection, the nature of what is involved in non-intellectually believing such a content succeeds. Under the non-intellectual way of having a moral or evaluative belief – be the belief true or false, be the believer ideally or less than ideally situated – one is bound to be attracted in some degree or other towards the option in question.

III. CONCLUSION

This concludes our presentation of moral functionalism and our explanation of how it can explain the linkage between evaluation and desire. The presentation of the functionalism, in the first part, was meant to meet the challenge associated with the Humean doctrine of an is–ought gap, and the explanation of the evaluation–desire linkage was designed to meet the challenge associated with the distinct Humean assumption that beliefs are motivationally inert.

The functionalism presented shows how a cognitive content can deliberately support a certain course of action, and how it can in this sense be evaluative. But it does so in a way that ought to be acceptable to Humeans and non-Humeans alike. The ‘is’ from which ‘ought’ is deriv-

able under our moral functionalism is a very special sort of 'is', since it bears on what the agent would desire under conditions in which certain recognized obstacles are absent, full information is available, and so on. The explanation of how evaluation connects non-inductively with desire ought also to be acceptable to Humeans and non-Humeans alike. It is certainly faithful to Hume in denying that there are any contents such that a belief in them necessarily involves the presence of desire. What it argues is only that there are certain contents such that the ordinary, canonical way of believing them involves desire.

We hope that our moral functionalism will prove to have appeal, though we readily acknowledge that we have not filled in the details; that we have done very little by way of meeting possible objections; and that we have done nothing to dislodge any of the alternative doctrines that currently do battle in the meta-ethical field. Our aim has been to float the functionalist idea, not to establish definitively that it is seaworthy.²

Research School of Social Sciences, Australian National University

REFERENCES

- Blackburn, S. 1984: *Spreading the Word* (Oxford UP).
 — 1993: *Essays in Quasi-Realism* (Oxford UP).
 Brink, D. 1986: 'Externalist Moral Realism', *Southern Journal of Philosophy*, Supp. Vol. 24, pp. 23–41.
 Dreier, J. 1990: 'Internalism and Relativism', *Ethics* 101, pp. 6–26.
 Gibbard, A. 1990: *Wise Choices, Apt Feelings* (Harvard UP).
 Hurley, S. 1989: *Natural Reasons: Personality and Polity* (Oxford UP).
 Jackson, F. 1992: Critical Notice of Susan Hurley's *Natural Reasons*, *Australasian Journal of Philosophy* 70, pp. 475–87.
 — 1994: 'Metaphysics by Possible Cases', *Monist* 77, pp. 93–110.
 Jackson, F. and Pettit, P. 1988: 'Functionalism and Broad Content', *Mind* 97, pp. 381–400.
 Lewis, D. 1983: *Philosophical Papers* (Oxford: Basil Blackwell).
 — 1989: 'Dispositional Theories of Value', *Proceedings of the Aristotelian Society*, Supp. Vol. 63, pp. 113–37.
 McDowell, J. 1978: 'Are Moral Requirements Hypothetical Imperatives?', *Proceedings of the Aristotelian Society*, Supp. Vol. 52, pp. 13–29.
 — 1979: 'Virtue and Reason', *Monist* 62, pp. 331–50.
 — 1981: 'Non-cognitivism and Rule-following', in S. Holtzman and C. Leich (eds), *Wittgenstein: to Follow a Rule* (London: Routledge).
 McNaughton, D. 1988: *Moral Vision: an Introduction to Ethics* (Oxford: Basil Blackwell).
 Pettit, P. 1991: 'Realism and Response-dependence', *Mind* 100, pp. 587–626.

² We owe special thanks to Michael Smith for discussion of the issues in the paper and for comments on earlier drafts. We are also very grateful to Simon Blackburn for an extended response to an earlier draft. We owe thanks for helpful comments to Geoff Brennan, Keith Campbell, Peter Menzies, Graham Oppy, Geoffrey Sayre McCord and anonymous referees for *The Philosophical Quarterly*.

- 1993: *The Common Mind: an Essay on Psychology, Society and Politics* (Oxford UP).
- Pettit, P. and Smith, M. 1993: 'Practical Unreason', *Mind* 102, pp. 53–80.
- Price, H. 1988: *Facts and the Function of Truth* (Oxford: Basil Blackwell).
- Quine, W.V.O. 1953: *From a Logical Point of View* (Harvard UP).
- 1974: *The Roots of Reference* (La Salle: Open Court).
- Rawls, J. 1971: *A Theory of Justice* (Oxford UP).
- Scanlon, T.M. 1982: 'Contractualism and Utilitarianism', in A. Sen and B. Williams (eds), *Utilitarianism and Beyond* (Cambridge UP).
- Smith, M. 1989 'Dispositional Theories of Value', *Proceedings of the Aristotelian Society*, Supp. Vol. 63, pp. 89–111.
- 1992: 'Valuing: Desiring or Believing?', in D. Charles and K. Lennon (eds), *Reduction, Explanation and Realism* (Oxford UP), pp. 323–60.
- Stocker, M. 1979: 'Desiring the Bad: An Essay in Moral Psychology', *Journal of Philosophy* 76, pp. 738–53.