

# DISPOSITIONAL THEORIES OF VALUE\*

Michael Smith, David Lewis and Mark Johnston

I—*Michael Smith*

The following propositions seem both plausible in their own right and apparently inconsistent:

(1) Moral judgements like ‘It is right that I  $\Phi$ ’ (‘valuations’ for short) express *beliefs*; in this case, a belief about the rightness of my  $\Phi$ -ing.

(2) There is some sort of a *necessary connection* between being in the state the judgement ‘It is right that I  $\Phi$ ’ expresses and having a motivating reason, not necessarily overriding, to  $\Phi$ .

(3) Motivating reasons are constituted, *inter alia*, by *desires*. The apparent inconsistency can be brought out as follows. From (1), the state expressed by a valuation is a belief, which, from (2), is necessarily connected in some way with having a motivating reason; that is, from (3), with having a desire. So (1), (2) and (3) together entail that there is some sort of necessary connection between distinct existences: a certain kind of belief and a certain kind of desire. But there is no such connection. Believing some state of the world obtains is one thing, what I desire to do in the light of that belief is quite another. Therefore we have to reject at least one of (1), (2) or (3). Call this the ‘moral problem’, and call those who respond ‘revisionists’ and ‘reconciliationists’.<sup>1</sup>

Revisionists accept the inconsistency, and so seek to explain away the apparent plausibility of at least one of (1), (2) and (3). Thus, for example, emotivists, prescriptivists and projectivists

\*The three papers that comprise this symposium come out of very extensive discussion between the three symposiasts—so extensive that sometimes we can no longer tell which ideas began with whom. But as will be plain, we have not ended in full agreement. Besides our indebtedness to one another, we are indebted also to Simon Blackburn, Paul Boghossian, Geoffrey Brennan, John P. Burgess, John Campbell, Josh Cohen, James Dreier, Jim Klagge, Robert Pargetter, Philip Pettit, Charles Pigden, Joe Pitt, Peter Railton, Denis Robinson, Gideon Rosen, Thomas Scanlon Jr., Nick Smith, Bas van Fraassen, Jay Wallace and Crispin Wright.

<sup>1</sup>The *locus classicus* for this form of argument is David Hume *Treatise of Human Nature* edited by Selby-Bigge (Oxford, 1888) pp. 413–7. See also John McDowell ‘Are Moral Requirements Hypothetical Imperatives?’ *Proceedings of the Aristotelian Society Supplementary Volume* 1978 p. 18; David McNaughton *Moral Vision* (Blackwell, 1988) p. 23.

accept (2) and (3) and reject (1).<sup>2</sup> Since we believe (1) only because sentences with moral contents display the syntactic and normative features of judgements that are truth-assessable—they figure in negation, conditional, propositional attitude contexts and the like, and particular uses of such sentences can be more or less justified—such theorists try to show how, via a logic of imperatives and universalizability, as in R. M. Hare's case, or 'quasi-realism', as in Simon Blackburn's, judgements that are not expressions of beliefs may still display these features. Alternatively, naturalists like Philippa Foot—at least when she wrote 'Morality as a System of Hypothetical Imperatives'—and Peter Railton, and contractualists like Thomas Scanlon accept (1) and (3) and therefore reject (2).<sup>3</sup> They argue that such evidence as there is for (2) is equally well explained by the fact that wide-spread moral education produces in many people not just moral beliefs, but also, entirely contingently, a desire to do what they believe they should. Finally, anti-Humean moral realists like Mark Platts and John McDowell accept (1) and (2), and therefore reject (3) (or perhaps they only accept (2), and reject both (1) and (3)).<sup>4</sup> They argue that the Humean philosophy of mind implicit in (3) (and perhaps also (1)) is indefensible; that, properly understood, motivating reasons may be constituted by cognitive states alone.

By contrast reconciliationists deny the inconsistency. Prominent amongst reconciliationists are those who accept a dispositional theory of value, a theory that purports to analyse value in terms of a disposition to take a favourable psychological attitude towards certain actions or outcomes under suitable

<sup>2</sup> A. J. Ayer *Language, Truth and Logic* (Blackwell, 1937) Chapter 6; R. M. Hare *The Language of Morals* (Oxford University Press, 1952); Simon Blackburn *Spreading the Word* (Oxford University Press, 1984) Chapter 6.

<sup>3</sup> Philippa Foot 'Morality as a System of Hypothetical Imperatives' in her *Virtues and Vices* (University of California Press, 1978); Peter Railton 'Moral Realism' *Philosophical Review* 1986; Thomas Scanlon 'Contractualism and Utilitarianism' in *Utilitarianism and Beyond* edited by Amartya Sen and Bernard Williams (Cambridge University Press, 1982). It should not be surprising to find Scanlon's contractualism in the company of Railton's consequentialist naturalism, for Scanlon explicitly formulates contractualism in opposition to such theories, versions of what he calls 'philosophical utilitarianism' as opposed to 'normative utilitarianism' ('Contractualism and Utilitarianism' pp. 108–10).

<sup>4</sup> John McDowell 'Are Moral Requirements Hypothetical Imperatives?'; Mark Platts *Ways of Meaning* (Routledge and Kegan Paul, 1979) Chapter 10.

conditions.<sup>5</sup> Such theorists attempt to explain why (1) is true by giving us an account of what rightness as an object of belief is, and why (2) is true, and consistent with (1) and (3), by treating 'taking a favourable psychological attitude' to be a species of desiring, thus displaying an internal, indeed *analytic*, connection between believing that  $\Phi$ -ing is right and desiring to  $\Phi$ . Importantly, a dispositional theorist may or may not claim that some valuations are true: that is, he may offer us an error theory in the spirit of John Mackie, a reconciliation at the level of concepts rather than ontology.<sup>6</sup> We will return to this point later.

If, as I believe, we really should accept the Humean philosophy of mind implicit in (3) (and (1)), then we must choose between reconciling, revising by rejecting (1), and revising by rejecting (2).<sup>7</sup> But how should we choose? Here we note a curious asymmetry. For while those who recommend revising by rejecting (1) seem to agree that reconciliation would be preferable—agree that (1) is, as it were, the *default* explanation

<sup>5</sup>David Lewis defends such a view in his contribution to this symposium. However, whereas I am concerned to give a dispositional theory of rightness, Lewis is concerned to analyse what it is for something to be a value quite generally. These theories are not in competition. A dispositional theory of rightness is best thought of as derived from, *inter alia*, a dispositional theory of value. If we derived such a theory of rightness from Lewis's theory of value, would the theory be reconciliationist in my sense? I am not sure. My reservations have to do with the kind of 'internalism' Lewis's theory secures. Views related to the dispositional theory of value are defended by C. D. Broad 'Some Reflections on Moral Sense Theories in Ethics' in *Broad's Critical Essays in Moral Philosophy* edited by David Cheney (Allen and Unwin, 1971); Roderick Firth 'Ethical Absolutism and the Ideal Observer' in *Philosophy and Phenomenological Research* 1952; John McDowell 'Values and Secondary Qualities' in *Morality and Objectivity* edited by Ted Honderich (Routledge and Kegan Paul, 1985); David Wiggins 'A Sensible Subjectivism' in his *Needs, Values, Truth* (Blackwell, 1987). For critical discussion see Jonathan Dancy and Christopher Hookway 'Two Conceptions of Moral Realism' *Proceedings of the Aristotelian Society Supplementary Volume* 1986; Crispin Wright 'Moral Values, Projection and Secondary Qualities' *Proceedings of the Aristotelian Society Supplementary Volume* 1988. Lewis is right, I think, not to make the case for the dispositional theory hang on the plausibility of a tight analogy between values and secondary qualities. I argue this point in my 'Objectivity and Moral Realism' in *Realism and Reason* edited by Crispin Wright and John Haldane (Oxford University Press, forthcoming).

<sup>6</sup>Mackie *Ethics: Inventing Right and Wrong* (Penguin, 1977) Chapter One. See also my 'Objectivity and Moral Realism'.

<sup>7</sup>For a discussion of (3) see John McDowell 'Are Moral Requirements Hypothetical Imperatives?'; my 'The Humean Theory of Motivation' *Mind* 1987; Philip Pettit 'Humeans, Anti-Humeans and Motivation' *Mind* 1987; my 'On Humeans, Anti-Humeans and Motivation: A Reply to Pettit' *Mind* 1988; David Lewis 'Desire as Belief' *Mind* 1988; John Collins 'Belief, Desire and Revision' *Mind* 1988.

of the syntactic and normative features of moral discourse<sup>8</sup>—many of those who recommend revising by rejecting (2) make no such concession. They deny even the *apparent* plausibility of (2).<sup>9</sup>

This seems to me to be a serious mistake and it is the task of this paper to say why. Such theorists say we should accept (1) but reject (2). But reflection on (1) itself reveals a reason for accepting (2): specifically, it reveals that we should accept (2) if we think that our reasons for our moral beliefs are themselves reasons for being morally motivated. I will argue that we should so conceive of our reasons for our moral beliefs and that, therefore, reconciliation is the preferred solution to the moral problem. Though I will not argue the point at length, it will emerge that those who opt for a dispositional theory of value are uniquely placed to explain these relations between our reasons for our moral beliefs and our reasons for our moral motivations. I close with some general remarks about the prospects for revision.

Let's begin by seeing whether there are any *independent* reasons for accepting (2); reasons that will remain intact even if we ultimately opt to revise by rejecting (1). Someone who thinks there are is doubtless impressed by the fact that the alleged connection between moral judgement and motivation is not an isolated feature of *moral* thought. For just think of what is involved, quite generally, in making evaluative judgements that are practical in their subject matter (from now on I will take this qualification as read).<sup>10</sup> Evaluative judgements are judgements that play a constitutive role in deliberation. For when we deliberate we make judgements about the *prima facie* desirability of our options and, on their basis, reach the conclusion a particular option is desirable all things considered.<sup>11</sup> Moreover,

<sup>8</sup> Consider, for example, Ayer's argumentative strategy described in my 'Should We Believe In Emotivism?' in *Fact, Science and Morality: Essays on A. J. Ayer's Language, Truth and Logic* edited by Graham Macdonald and Crispin Wright (Blackwell, 1987).

<sup>9</sup> The exception is Scanlon 'Contractualism and Utilitarianism' pp. 117–8.

<sup>10</sup> This seems to be Tom Nagel's strategy in defence of the requirement in his *The Possibility of Altruism* (Oxford University Press, 1970) p. 9. Note that in what follows I will sometimes refer to the state that is expressed by an evaluative judgement as an 'evaluative thought'. Here too I will take as read the qualification 'that is practice! in its subject matter'.

<sup>11</sup> Are these judgements of 'prima facie' or 'pro tanto' desirability? See Donald Davidson 'How is Weakness of Will Possible?' in his *Essays on Actions and Events* (Oxford University Press, p. 198); Susan Hurley 'Conflict, Akrasia and Cognitivism' *Proceedings of the Aristotelian Society* 1985–6.

we reach this conclusion about a particular option in order to do it. But how would this be possible if there were no connection between judgements of prima facie desirability and desiring? What *further* consideration could possibly get us motivated to do what we judge prima facie desirable if we could coherently think 'Yes, this option is prima facie desirable, but so what? I don't *care* whether my actions are prima facie desirable or not'?

It might be suggested that what's needed is a *desire* to do what's prima facie desirable. But this suggestion has to be made to cohere with the fact that it is *by deliberating* that we come to act in the way that we do, so the question is 'How is that desire connected with deliberating?' and no plausible answer consistent with the present line of objection seems to be forthcoming. For that desire would have to be connected with some *further* evaluative judgement, say, that it is prima facie desirable to do what's prima facie desirable. And then we would have to ask why *that* judgement is supposed to guarantee motivation if the original judgement that a certain option is prima facie desirable doesn't. A desire to do what's prima facie desirable is needed alright, but that is just to concede the point. For all that shows is that for deliberation to be possible at all there must be some sort of necessary connection between being in the state the evaluative judgement ' $\Phi$ -ing is prima facie desirable' expresses and desiring to  $\Phi$ . And, of course, that is just to say that (2) may be reconstrued as a quite general constraint on evaluative judgements (call this ' $2^c$ ').

Given that this is so it might be said that the task of defending (2) is made much easier. For it is up to someone who challenges (2) to say why moral thought *doesn't* have this feature. How surprising that would be given that moral thought is simply a *species* of evaluative thought!

Moreover, it might be said, we can now see that the moral problem doesn't trade on particular features of *moral* thought. Rather it is an instance of a quite general problem concerning evaluative thought. For, recall that the idea was that we deliberate on the basis of judgements concerning the prima facie desirability of our options. The general problem can be put like this: 'How can *any* evaluative judgement be what it seems to be, the expression of a belief about the desirability of an option, given that it has to satisfy ( $2^c$ )? After all, belief and desire are

distinct existences and. . .'. We may call this the 'deliberative problem', and, as before, we can divide those who respond into reconciliationists and revisionists.<sup>12</sup> The deliberative problem has attracted much attention. Significantly, however, no one who recommends revision in response to the deliberative problem seriously suggests that we should reject (2<sup>c</sup>). No surprise given that that would make deliberation practical only in its subject matter, not in its issue. If we are right to see the moral problem as a mere instance of the deliberative problem then, by parity of reasoning, no-one should seriously suggest rejection of (2) in response to the moral problem either.

Finally, we can now state more clearly what (2) says. For the argument just given establishes that there is, at least, a *defeasible* connection between evaluative judgements and the will; that in the normal case no further judgement, and thus no further desire, is required to get someone who judges that his  $\Phi$ -ing is *prima facie* desirable to desire to  $\Phi$ . But what does this idea of the 'normal case' amount to? This question is explored in some detail by Michael Stocker who observes:

Through spiritual or physical tiredness, through accidie, through weakness of body, through illness, through general apathy, through despair, through inability to concentrate, through a feeling of uselessness or futility, and so on, one may feel less and less motivated to seek what is good. One's lessened desire need not signal, much less be the product of, the fact that, or one's belief that, there is less good to be obtained or produced, as in the case of a universal Weltschmerz. Indeed, a frequent added defect of being in such 'depressions' is that one sees all the good to be won or saved and one lacks the will, interest, desire or strength.<sup>13</sup>

What Stocker's discussion suggests is that a subject is motivated to do what he judges *prima facie* desirable *only if nature conspires*;

<sup>12</sup> Anthony Kenny *Action, Emotion and Will* (Routledge and Kegan Paul, 1963) and Hector Neri Castañeda *Thinking and Doing* (Reidel, 1975) presumably count as revisionists. Donald Davidson in 'How is Weakness of the Will Possible?' seems to be a reconciliationist, but see p. 86 of 'Intending' in his *Essay on Actions and Events*.

<sup>13</sup> Stocker 'Desiring the Bad: An Essay in Moral Psychology' *Journal of Philosophy* 1979 p. 744.

only if one of these ‘depressions’ does not interfere with that connection. If to be depressed in the relevant way is to be subject to a condition of psychological ill-health then the more general idea at work here is plausibly that if A judges his  $\Phi$ -ing to be prima facie desirable then either he has a motivating reason to  $\Phi$  or he is *irrational*. If this is right, (2) should be understood accordingly.<sup>14</sup>

We have, then, a clearer idea of what (2) says, and, given the assumption that moral thought is a species of evaluative thought, we have established a presumption in its favour. For (2) is an instance of (2<sup>c</sup>), a quite general constraint on evaluative judgements. How might someone challenge (2), given this defence? As we shall see, a considerable challenge emerges as a corollary to Philippa Foot’s attack on rationalism in her ‘Morality as a System of Hypothetical Imperatives’.

According to Kant, moral requirements are categorical, not hypothetical, imperatives. Moreover, Kant thinks it follows from this that moral requirements are requirements of reason. However, Foot argues, Kant faces a dilemma. On the first horn, though we can agree that moral requirements are expressed by judgements employing a categorical use of ‘should’, this does not entail that moral requirements are requirements of reason. On the other horn, Foot asks why, in that case, Kant thinks moral requirements *are* requirements of reason, insisting that his claim needs a special kind of support, a kind of support that doesn’t seem to be forthcoming once we fully appreciate the arguments on the first horn. Let me briefly explain her argument on each side of this dilemma.

<sup>14</sup> It should be said that Stocker considers and rejects the claim that the gap between evaluation and motivation is evidence of irrationality. However his reason for rejecting it seems to me based on a confusion. For he rejects it simply because not *all* manifestations of despair, accidie, apathy and the like are themselves evidence of irrationality. But though we might agree that *some* such manifestations are not evidence of irrationality, the question is whether we have such evidence when they manifest themselves by realising their potential to interfere with the connection between evaluation and motivation. Myself I think that we do. In such cases subjects are quite uncontroversially irrational in at least this sense: *the psychological processes of the perfectly rational agent are subject to no such interference*. This claim seems to me independently plausible. But, if it seems in need of support, note that the psychological failures Stocker mentions are quite generally apt for interfering with rational connections between psychological states. We characteristically appeal to such failures when we seek to explain why people fail to perform rational inferences, for example.

Foot begins by explaining why everyone should accept that moral judgements are expressed by judgements employing a categorical 'should'. Suppose a man wants to go home and we tell him that he should catch the noon train, but that just before noon he stops wanting to go home. We then have to withdraw the claim that he should catch the noon train. Here, then, we have a hypothetical 'should', for the truth of the 'should' claim is conditional on what the man wants. Contrast the case where a man behaves in a way contrary to some moral requirement—suppose he is cruel. It remains true to say that he behaved as he shouldn't even if he tells us (truly) that not being cruel in these circumstances would in no way serve any interest or desire of *his*. Here, by contrast, we have a categorical use of 'should'. The difference marked by the categorical, as opposed to the hypothetical, 'should' thus has to do with the kinds of consideration required to support the 'should' claim, the difference between those 'should' claims that can only be supported by showing how the action in question serves a desire or interest of the agent (the hypothetical) and those where this is not so, where what is required is mention of some relevant feature of the agent's circumstances (the categorical).

However, as Foot reminds us, in this sense moral requirements, though categorical, are like many other more mundane requirements: for example, requirements of etiquette. Suppose someone acts contrary to a requirement of etiquette—he replies in the first person to a letter written to him in the third. It remains true that he acted as he shouldn't even if he tells us (truly) that acting as etiquette requires in no way serves any interest or desire of his. For, as with requirements of morality, what *makes* it true that some action is required by etiquette *isn't* that acting in the relevant way serves an interest or desire of the agent.

The fact that requirements of etiquette and morality seem in this respect to be on all fours provides Foot with the materials to complete her argument on the first horn of the dilemma and make her argument on the other horn. For no-one holds that requirements of etiquette are requirements of reason; that someone who replies in the first person to a letter written to him in the third is shown thereby to be *irrational!* But in that case we have an example of a requirement appropriately expressed by a



categorical use of 'should' that is not itself a requirement of reason. And then Foot argues, on the other horn of the dilemma, that someone who insists that moral requirements *are* requirements of reason had better tell us why we shouldn't think that moral requirements are on all fours with requirements of etiquette in this respect as well. He must tell us in some non-question-begging way what the relevant *difference* is supposed to be between norms of morality and norms of etiquette.

We are now in a position to see why a challenge to (2) emerges as a corollary. For an agent's being disposed to make judgements about what *etiquette* requires of him tells us nothing about what he is disposed to *do*. That is, (2) is *certainly* unacceptable if 'right' is read as 'required by etiquette'.<sup>15</sup> Foot may thus quite legitimately challenge the defender of (2) to say why he thinks (2) is acceptable if 'right' is read as 'required by morality'. What is the relevant *difference* supposed to be between norms of etiquette and norms of morality?

Foot's challenge is considerable. For it forces us to make a distinction between an agent's *recognition* of a norm and his *embracing* that norm.<sup>16</sup> When I say 'A letter written in the third person should be answered in the third person' my judgement reflects my *recognition* of the relevant norm of etiquette. But in order to be moved by that consideration when I am writing a letter I have to do more than recognize the norm, I have to *embrace* it: that is, very roughly, I have to want to answer letters in the way I know I should. And this in turn shows what was inadequate about the defence of (2) given earlier. For what we saw then was that, in these terms, there is at least *one* category of thought with normative contents where the distinction between recognising and embracing a norm is *collapsed*: namely, evaluative thought. However, though in the case of evaluative thought that distinction is quite rightly collapsed—otherwise evaluative judgements would not be able to play the role they play in deliberation—Foot challenges us to say why we simply *assume* that moral judgement is, of its nature, a mode of expressing evaluative thought. Why not think instead that moral thought is

<sup>15</sup> In the words of the children's song, 'You may very well be well-bred/Lots of etiquette in your head/But there's always some special case/Time or place, to forget etiquette . . . for instance/Never smile at a crocodile. . .'

<sup>16</sup> See Railton 'Moral Realism' p. 168.

simply, like thoughts about the requirements of etiquette, *recognitional*, with *embracing* the requirement in question being an optional extra?

It might be thought that the defender of (2) shouldn't be too worried by this challenge; that he can insist that, since Foot says nothing to show that we have to *deny* (2), so her challenge shows at most that if we do accept it, we accept it as *basic*. However it seems to me that this would be bad enough. For those who reject (2) might equally reply that, if we deny (2), we take its *denial* to be basic. There would then be a real question what interest *debates* between those who respond to the moral problem by accepting (2) and those who respond by rejecting it should hold for us. For neither would be able to say anything in favour of especially *his* attitude towards (2). It is time to examine the plausibility of the assumption that our reasons for accepting (1) and (2) are entirely independent of each other.

We have already seen that reconciliationists and revisionists can agree that we have some reason to believe (1). For they can agree that belief is the appropriate attitude to contents expressed in sentences that are truth-assessable, and they can agree that sentences with moral contents display the syntactic and normative features of such sentences. I want now to argue that they can agree about more. For a moment's reflection reveals something about the *kind of justification* a moral belief seems to require, and thus something about the *kind of content* a moral judgement seems to have. In the end I want to argue that the kind of belief we have reason to believe a moral belief is, suggests a reason for accepting (2).

The point I have in mind about moral belief emerges as soon as we think about moral arguments. Suppose A says 'Φ-ing is worthwhile' and B says 'Φ-ing is not worthwhile'. If the value in question is moral value, then we seem immediately to conclude that at least one of A or B is *mistaken*. *Argument* ensues: 'What are your reasons for believing that Φ-ing is worthwhile? Display them to me so that I too can see the value of Φ-ing'. If such reasons aren't forthcoming then A may rightly assume that B doesn't have adequate reasons to believe that Φ-ing is worthwhile.

Indeed, this seems to me to be partially constitutive of moral value, as against values of other kinds. Suppose A and B disagree over some aesthetic matter: the relative merits of, say, Turner

and Pollock. In this case we seem much happier to rest content with bafflement at why someone likes what we can't stand; much happier to admit that, since 'there's no accounting for taste', so, at bottom, we have a *mere* difference in taste. Given that there is no reason to *criticize* A for judging Pollock to be a better painter than Turner, and vice versa, we might explain away the apparent conflict in these judgements by providing a relativistic analysis of them: A's judgement is about Turner's and Pollock's respective standings relative to *his* aesthetic sensibility whereas B's is about their standings relative to *his*. We are, however, profoundly resistant to resting content with such bafflement in the case of moral value. Moral difference about the relative importance of justice and self-interest in a particular case, say, seems not to be explicable in terms of a mere difference in *taste*. And thus we are profoundly resistant to analysing away moral disagreements by finding a hidden relativity in our moral judgements.<sup>17</sup>

Nor should it be thought that this is a mere artefact of *moral argument*; that since in moral argument our task is to resolve conflicts, so we keep looking for common ground, but that this is consistent with our accepting that, at bottom, there may be no common ground to be found.<sup>18</sup> For these points have nothing especially to do with moral arguments. Rather, so long as we can imagine some *hypothetical* rational creature to whom we cannot justify our moral beliefs, the search for reasons in support of them is in place. (Just think of the method of moral theorising.) Thus, we may say: moral beliefs seem to be beliefs about some non-relative fact of the matter and the search for reasons in support of our moral beliefs seems to be the search for reasons that would convince rational creatures *as such* to take on such beliefs for themselves.

We are now in a position to see why reflection on the kind of belief a moral belief seems to be suggests a reason for accepting (2). For since, according to (2), someone who has a moral belief but is not motivated accordingly is, to that extent, *irrational*, so, if (2) is true, our moral motivations can be seen to *inherit* the

<sup>17</sup> See Bernard Williams *Ethics and the Limits of Philosophy* (Harvard University Press, 1985) pp. 158–9.

<sup>18</sup> This line was suggested to me by conversations with Simon Blackburn.

rational status of our moral beliefs. Now what we have just seen is that we seem to conceive of our moral beliefs as beliefs that require a *privileged rational defence*. If (2) is true, then we should presumably conceive of our moral motivations as motivations that *themselves* require a privileged rational defence. But do we? I want to argue that we do. Moreover, I want to argue that the best account of what it is to conceive of our moral motivations as requiring a privileged rational defence suggests that our reasons for being morally motivated are *themselves* our reasons for our moral beliefs. Accordingly, it seems to me, we should accept both (1) and (2).

Foot would doubtless be unimpressed. For the idea that our moral motivations require a privileged rational defence *just is* the idea that moral requirements are requirements of reason. Yet she began by asking the rationalist what reason we have for supposing that to be true. Let's therefore focus on Foot's challenge to the rationalist.

Foot wants the rationalist to tell us, in a non-question-begging way, why we should think that moral requirements are requirements of reason. But this is ambiguous. Is he to justify the claim that we *conceive* of moral requirements as requirements of reason—something that allows that we may be wrong in supposing that *there are* any moral requirements—or the claim that, in addition, *there are* some such moral requirements? Clearly, the argument for (2) requires only that he be able to do the first, not the second. For (2) simply tells us, *inter alia*, that we must think of our moral beliefs and motivations as alike in rational status. Thus all that's required is that, as we saw with regard to our moral beliefs, our moral motivations *seem* to require a privileged rational defence; that we *conceive* of them as requirements of reason.<sup>19</sup>

At the end of 'Morality as a System of Hypothetical

<sup>19</sup> Foot does not keep these different claims separate. Thus at one point she urges her challenge against the rationalist by insisting 'Irrational actions are those in which a man in some way defeats his own purposes, doing what is calculated to be disadvantageous or to frustrate his ends. Immorality does not *necessarily* involve any such thing' ('Morality as a System of Hypothetical Imperatives' p. 162). But whether or not we think this a good argument, it is irrelevant if our interest is in whether we *conceive* of moral requirements as requirements of reason. Is it a good argument? That depends on whether there are principles of reason like the principle of limited altruism (see later).

Imperatives' Foot seems to acknowledge an explicit argument for this claim. She remarks:

We are apt to panic at the thought that we ourselves, or other people, might stop caring about the things we care about, and we feel that the categorical imperative gives us some control over the situation.<sup>20</sup>

If anywhere, she seems to be saying, the fact that we conceive of moral requirements as requirements of reason manifests itself in circumstances like this; manifests itself in the apparent *legitimacy* of this kind of panic. In the end, however, she concludes that our conception of moral requirements does not allow us to make sense of this sort of panic. For, she argues, when we think more carefully about moral concern we see that it is unnecessary to have the kind of control we might think the categorical imperative gives us. Moreover, she argues, this is just as well, for the idea that the categorical imperative gives us such control is an 'illusion'; it is 'as if we are 'trying to give the moral "ought" a magic force'.<sup>21</sup> Let me consider these claims in turn.

Foot argues that it is unnecessary to have the control we might think the categorical imperative gives us because, in other areas of life, we get by without it. Thus, she observes,

. . . it is interesting that the people of Leningrad were not struck by the thought that only the *contingent* fact that other citizens shared their loyalty and devotion to the city stood between them and the Germans during the terrible years of the siege.<sup>22</sup>

*They* did not panic at the thought that loving Leningrad is not rationally required. By analogy, then, Foot asks: 'What is wrong with thinking of the moral person as someone who simply loves the ends that morality aims at: liberty, justice, truth and the like? What is wrong with thinking of a community of moral agents as being like "volunteers banded together to fight for liberty and

<sup>20</sup> Foot 'Morality as a System of Hypothetical Imperatives' p. 167.

<sup>21</sup> Foot 'Morality as a System of Hypothetical Imperatives' p. 167.

<sup>22</sup> Foot 'Morality as a System of Hypothetical Imperatives' p. 167. Note that Foot once again focusses on the non-moral and asks what the relevant difference is with the moral. (Like Foot, I will ignore the fact that the people of Leningrad were no doubt actually *morally* motivated.)

truth, and against inhumanity and oppression”?’<sup>23</sup> Foot can see nothing wrong with this. And so she rejects the idea that the panic she describes is even apparently legitimate.

Foot’s question is a significant one, but I suspect that an answer to it pushes us in the opposite direction. If the citizens of Leningrad were asked why they bothered fighting for their city then they would presumably have begun by listing all of the things about their city that they loved. But if they were asked why *those* things mattered to them, and not other things—things about Berlin for example—then they would presumably have simply rested content with the fact that they could give no distinctively rational justification of their caring about these things; rested content with the fact that, since they grew up in Leningrad not Berlin, so *these* are the things that they love, not *those*.

But now suppose we asked someone why he bothered fighting for liberty and justice. Is the corresponding story plausible? I do not think so. In order to see this, note, to begin, that as we have seen, moral agents do not take the objects of their moral concern to be fixed in the way the objects of our love and devotion are fixed. Thus we think that we might be *wrong* to assign the moral value we do assign to liberty and justice. Is this the same as thinking we might be *wrong* that we love liberty and justice? No. The thought is rather that liberty and justice may not be deserving of the love and devotion we do *in fact* give them. It therefore seems that, in defending fighting for liberty and justice, we would *not* simply cite the features of liberty and justice we love—as the citizens of Leningrad might rightly have cited the features of Leningrad they loved—but would rather try to give our reasons for thinking that these features of liberty and justice *merit* our love and devotion. The question whether we conceive of moral requirements as requirements of reason then turns on how we should interpret this idea of *meriting* love and devotion. Do liberty and justice merit our love and devotion only if the considerations we cite in their support have appeal, not just to *ourselves*, but to any creature capable of asking the question ‘Should I care about liberty and justice?’, at least insofar as they are rational? If so then, *a fortiori*, we will not rest

<sup>23</sup> Foot ‘Morality as a System of Hypothetical Imperatives’ p. 167.

content with saying that these are things that we *in fact* care for. But how are we to determine whether that is how we should interpret the idea of meriting love and devotion?

Foot is right that, if it is, then this will manifest itself in the apparent legitimacy of a certain sort of panic. But it is panic not at the thought of the *contingency* of our caring about the things we care about, but rather at the thought that if we stopped caring, or if we had never started, we would rightly come to see, or perhaps always have seen, our caring about these things as *rationally optional*; as *arbitrary*; as to be *explained* rather than *justified*; as to be explained, in much the same way that loving Leningrad is to be explained, simply in terms of *when, where and how we were brought up*.<sup>24</sup> That seems to me to be a kind of panic that we rightly feel when we reflect upon the possibility that we can give no privileged rational defence of moral concern and that the citizens of Leningrad rightly did not feel when they reflected on the fact that they could give no privileged rational defence of loving Leningrad.<sup>25</sup>

In order to see this, suppose I care for justice as well as for myself and you only care for yourself. Suppose you make your case for caring only for yourself and I make my case for caring for justice as well. Suppose we fail to convince each other and begin wondering whether this is a difference between us to be explained rather than justified. We rightly panic if we take this idea seriously because, I suggest, we can then no longer take seriously, as we must, the idea of *disapproving* of someone for caring other than the way we think he should.<sup>26</sup>

Imagine a citizen of Leningrad who claimed to *disapprove* of a Berliner for loving features of Berlin rather than Leningrad. Can we make sense of this attitude? No. The attitude seems altogether inappropriate because we know that if you grow up in Berlin you will love Berlin, and if you grow up in Leningrad you will love Leningrad. Having grown up in Leningrad you

<sup>24</sup> Thus, though it may be a *contingent* matter whether I care about justice, it may not be *arbitrary*. Caring about justice may be rationally required, and so not arbitrary, despite the fact that, since it is a contingent fact that I am rational, so I only contingently care about justice.

<sup>25</sup> The thought that requirements of etiquette are arbitrary, in the relevant respect, inspires no panic either.

<sup>26</sup> Here I find myself in agreement with much of Foot's 'Approval and Disapproval' in her *Virtues and Vices*, though it seems that I take her argument to have implications that she would herself reject.

have every reason not to want to live in Berlin, and vice versa. But the idea that there is *something wrong* with someone from Berlin who doesn't want to live in Leningrad, that he makes some kind of *mistake*, is simply absurd. *Disapproval* is out of place.

Of course, citizens of Leningrad may well *dislike* Berliners. But that is because, like loving Leningrad itself, what we like and dislike about people is, in a certain respect, *arbitrary*. Thus it is significant that we are not rationally constrained to dislike people who like what we don't. For disliking, unlike disapproving, of someone does not imply you think there is something wrong with him, that he makes some kind of mistake. Disliking someone is, as it were, as much a fact about you as a fact about him. Consequently, the fact you dislike certain features of someone I don't is a fact about you I don't have to take all that seriously. It need provide me with no reason to suppose I should dislike him.<sup>27</sup> And hence, from a certain perspective, the fact that I don't like certain features of someone is not a fact about myself I have to take all that seriously either. Thus it is significant that we can quite seriously entertain the idea of *undoing* our likes and dislikes. Consider a frivolous example. I dislike it when people constantly grin; I find it annoying. However I couldn't really care less whether I *continue* to find this feature of people annoying. Getting rid of my distaste for constant grinning is something I would readily consider if someone could give me a method and a motive for doing so, and, correspondingly, is something I will not consider if someone does not give me a method and a motive for doing so.

And now the explanation of why we panic at the thought that moral concern can be given no privileged rational defence can be put like this. Once I come to see the fact that I care for both justice and myself and you care only for yourself as a difference between us to be explained rather than justified, the idea of my *disapproving* of you for caring only for yourself begins to look as absurd as a citizen of Leningrad *disapproving* of a Berliner for loving Berlin; as absurd as my *disapproving* of someone who constantly grins. After all, there is nothing *wrong* with you for not

<sup>27</sup> Of course, it may provide me with a motive for not liking him. If I don't like him then perhaps that will jeopardize our friendship and I very much want our friendship to continue.



caring for justice; you make no *mistake*. At best I *dislike* the fact that you don't care for justice, just as you, perhaps, dislike the fact that I don't care just for myself. But what we like and dislike about people is *arbitrary*. I therefore needn't take seriously your disliking what you dislike about me, and you needn't take seriously my disliking what I dislike about you. Indeed, from a certain perspective, neither of us should take our own dislikes all that seriously either. Perhaps we should seriously consider *undoing* our dislikes. I wonder whether we have a motive for doing so. . . ? Panic seems quite rightly to have set in.

Note how unsatisfying it would be to be told at this point that in actually disliking those who don't care for justice, I also dislike myself in possible worlds in which I don't care for justice; that this is why I can't seriously consider undoing this particular dislike of mine; that this is the relevant difference with say, my distaste for constant grinning.<sup>28</sup> For even if this makes for a difference, it does nothing to remove the *arbitrariness* of that dislike, and it is the arbitrariness of the dislike that is the source of the unease. After all, in another world I dislike myself in this world for not caring only for myself. How *peculiar* each of these attitudes seems, in the context of the other! Far more plausible is the idea that my actual attitude towards myself in worlds in which I don't care for justice is itself dependent upon my belief that I am, in the actual world, possessed of a *justification* for caring for justice. For, it seems, only this could account for my apparent ability to question whether I should have the attitude I actually have towards myself in such worlds. Panic sets in when I entertain the thought that this is an ability I don't actually have.<sup>29</sup>

Foot thus seems to me wrong to suppose that it is unnecessary to have the 'control' the categorical imperative gives us. For the idea that moral concern is required by reason seems to be the only thing that would allow us to make sense of the idea of moral disapproval in this context, the idea that there is *something wrong* with someone who is *motivated* to do other than what he should;

<sup>28</sup> For suggestions along these lines see Martin Davies and Lloyd Humberstone 'Two Notions of Necessity' *Philosophical Studies* 1980 pp. 22-5; my 'Should We Believe in Emotivism?' pp. 303-4; David Wiggins 'A Sensible Subjectivism?' pp. 205-6.

<sup>29</sup> This may be seen as correcting what I say in 'Should We Believe in Emotivism?' pp. 303-4.

that he makes some kind of *mistake*. For, we may then say, what is wrong with him is the same as what is wrong with someone who *believes* other than what he should. *Each is insufficiently sensitive to the available reasons.*

This conclusion is certain to meet with resistance. For just as there are people who take the idea that moral requirements are requirements of reason to be axiomatic, there are others who treat the idea with evident disdain.<sup>30</sup> The following remarks of Simon Blackburn's are representative:

This is the permanent chimaera, the holy grail of moral philosophy, the knock-down argument that people who are nasty and unpleasant and motivated by the wrong things are above all *unreasonable*: that they can be proved to be wrong by the pure sword of reason. They aren't just selfish or thoughtless or malignant or imprudent, but are reasoning badly, or out of touch with the facts. It must be an occupational hazard of professional thinkers to want to reduce all the vices to this one. In reality the motivational grip of moral considerations is bound to depend on desires which must simply be taken for granted, although they can also be encouraged and fostered.<sup>31</sup>

However, the problem with this is that it doesn't speak to the issue.

If moral requirements are requirements of reason then we have no need to deny that someone who is motivated by the wrong things *is* just malignant; for being malignant *is* just a manifestation of an insensitivity to good reasons. The rationalist's idea is not that we need to *prop up* our terms of moral assessment with terms of rational assessment because the moral terms aren't enough by themselves. The idea is rather that, in order to understand *why* our terms of moral assessment are enough by themselves, we have to think of moral requirements as requirements of reason. To think otherwise is to suppose that the charge 'He is malignant!' is much like the charge 'He is from Berlin!', or 'He constantly grins!' or 'He answers letters written

<sup>30</sup> Tom Nagel takes the idea to be 'self-evident' in *The View From Nowhere* (Oxford University Press, 1986) pp. 159–60.

<sup>31</sup> Simon Blackburn *Spreading the Word* p. 222.

to him in the third person in the first!' And that is plainly wrong.<sup>32</sup>

Nor need the rationalist deny that that the motivational grip of moral considerations 'is bound to depend on desires which must simply be taken for granted'.<sup>33</sup> For it is now a familiar point that, at a certain level, at 'bedrock', our disposition to say, for example, '12' when asked 'What is  $7 + 5$ ?' depends on a disposition 'which must simply be taken for granted'.<sup>34</sup> But if no one needs to take a transcendental view of reason in order to see the rules of arithmetic and logic as justified, then why insist that the rationalist needs to take a transcendental view of reason in order to see the rules of morality as justified? *Bedrock* should not be thought *arbitrary*.<sup>35</sup>

Another reason why the foregoing argument might meet resistance is that, if what I have said is right, it follows that it is in fact *inappropriate* to say that someone committed to acting on requirements of etiquette *disapproves* of those who couldn't care less. But is this plausible? After all, it might be said, they would surely *say* they disapprove. But whether they would say this or not, it seems plausible to me to suppose that they do not really disapprove in the sense of supposing that there is *something wrong* with someone who couldn't care less about the requirements of etiquette; that he is making some sort of *mistake*. Indeed, though to my ear it sounds more appropriate to say that someone committed to acting on requirements of etiquette *dislikes* those who couldn't care less, even this seems to me a non-compulsory description of their attitude. For note that the dismissive remark 'He isn't one of us' has a legitimate role to play in practices of etiquette. Accordingly it seems to me that those who couldn't care less aren't so much thought to be mistaken or disliked by

<sup>32</sup> Thus I cannot agree with David Lewis when he writes in his contribution to this symposium 'And it won't help to juggle terms; as it might be, by calling it 'rational necessity' and then classifying the disposition to value genuine values as a department of 'rationality' (p. 133).

<sup>33</sup> Can a rationalist agree that motivating reasons are constituted *inter alia* by desires? Yes he can. See my 'Reason and Desire' *Proceedings of the Aristotelian Society* 1987-8, section III and IV.

<sup>34</sup> See, for example, Saul Kripke *Wittgenstein on Rules and Private Language* (Harvard University Press, 1982); Crispin Wright *Realism: Meaning and Truth* (Blackwell, 1986). See also Simon Blackburn *Spreading the Word* pp. 197-8.

<sup>35</sup> See Mark Johnston's contribution to this symposium.

those who could, as simply avoided and ignored. (Myself I find this rather alarming in certain contexts.)

Consider now Foot's second argument. The idea that the categorical imperative gives us the kind of control over our concerns that we want is, she suggests, an 'illusion'. It is 'as if' we are 'trying to give the moral "ought" a magic force'. What does Foot mean by this? As I understand it, she is questioning the *coherence* of the categorical imperative; asking how moral concern *could* be required by reason; asking for the *form* of an answer. But the answer is easily given.<sup>36</sup>

Consider something uncontroversially required by reason: that we desire the means to our ends (here we see the relevance of the earlier discussion of (Z<sup>c</sup>)). To say desiring the means to our ends is required by reason is to say, first, that since some version of the means-ends principle

(x)(If x's  $\Phi$ -ing is prima facie desirable and x can  $\Phi$  by  $\Psi$ -ing then x's  $\Psi$ -ing is prima facie desirable)

is valid<sup>37</sup>—i.e. a principle allowing us to derive evaluative truths from evaluative truths and truths about means—so this principle can be used to explain the *rationality* of deliberation.<sup>38</sup> For we can then see why a subject who believes both that his  $\Phi$ -ing is prima facie desirable and that he can  $\Phi$  by  $\Psi$ -ing but who doesn't believe that his  $\Psi$ -ing is prima facie desirable *violates* a norm of reason.<sup>39</sup> However it is also to say, second, that since a subject who believes his  $\Phi$ -ing is prima facie desirable is irrational if he does not desire to  $\Phi$ —something we learnt from

<sup>36</sup> Here I have been influenced by Christine Korsgaard 'Skepticism about Practical Reason' *Journal of Philosophy* 1986. For my disagreements with Korsgaard see sections III and IV of my 'Reason and Desire'. In what follows I assume that reconciliation is the *preferred* solution to the deliberative problem (legitimately—see the earlier discussion of (Z<sup>c</sup>) and footnote 38 below). I hope to defend such a solution elsewhere.

<sup>37</sup> The validity of this principle requires that the derived judgement be understood as a judgement of prima facie *instrumental* desirability. For other complications see my 'Reason and Desire' footnote 11. These complications may safely be ignored here however, for the argument I am constructing doesn't turn on the precise form of the principle, merely on its acceptability in some form or other.

<sup>38</sup> The assumption is that, as with moral judgements, we have reason to believe evaluative judgements express beliefs because of the syntactic and normative features of the sentences in which they are expressed. I hope to expand on this idea elsewhere.

<sup>39</sup> Compare someone who believes both that  $p$  and that  $p \rightarrow q$  but who does not believe that  $q$ . See my 'Reason and Desire' section II.

the earlier discussion of (2<sup>c</sup>)—so a subject who believes both that his  $\Phi$ -ing is prima facie desirable and that he can  $\Phi$  by  $\Psi$ -ing but who isn't motivated to  $\Psi$  also violates a norm of reason.

If this is right then to say moral concern is required by reason is to say something straightforwardly analogous. It is to say, first, that since a principle like the principle of limited altruism

(x) (If someone is in pain and  $x$  can relieve that person's pain by  $\Psi$ -ing then  $x$ 's  $\Psi$ -ing is prima facie desirable)

is valid—i.e. expresses a principle allowing us to derive evaluative truths from truths about our circumstances—so this principle can be used to explain the rationality of deliberation.<sup>40</sup> For we can then see why a subject who believes that someone is in pain and he can relieve that person's pain by  $\Psi$ -ing but who doesn't believe it is prima facie desirable that he  $\Psi$ 's violates a norm of reason. But it is also to say, second, that since a subject who believes his  $\Phi$ -ing is prima facie desirable is irrational if he does not desire to  $\Phi$ , so a subject who believes both that someone is in pain and that he can relieve that person's pain by  $\Psi$ -ing but who doesn't desire to  $\Psi$  also violates a norm of reason.

I thus see no problem with providing the *form* of an answer to the question 'How *could* it be that moral concern is required by reason?' Foot may well, of course, be right that it is an 'illusion' to suppose that moral concern is required by reason, for she may be right to be skeptical about the existence of principles like the principle of limited altruism allowing us to derive evaluative truths from truths about our circumstances. But if what I have said here is right, skepticism about the existence of such principles may well be tantamount to skepticism about the existence of moral requirements themselves.

Recall that, according to the argument for (2) sketched above, we should accept (2) if we think our moral motivations require a privileged rational defence and if the best account of what it is to so conceive of our moral motivations suggests that our reasons for being morally motivated are themselves our reasons for our moral beliefs. For then, by (2), our moral motivations may be seen to inherit the rational status of our

<sup>40</sup> See Tom Nagel *The View From Nowhere* (Oxford University Press, 1986) pp. 156–62. For a different way of putting the same point see my 'Reason and Desire' section IV.

moral beliefs. It should now be evident that the argument given establishes just this conclusion. For if principles like the principle of limited altruism are valid, then it is precisely these principles we will appeal to in justifying our moral beliefs. Moreover, as I said earlier, reconciliationists who opt for a dispositional theory of value are uniquely placed to explain why this is so.

According to the dispositional theorist we can analyse rightness in terms of a disposition to desire under suitable conditions. But what are these 'suitable conditions'? The preceding discussion provides an answer. The dispositional theorist should say that an agent's  $\Phi$ -ing is right just in case he would desire to  $\Phi$  if he were to deliberate in accordance with the principles of reason corresponding to moral principles, principles that permit us to derive evaluative truths from truths about our circumstances, principles like the principle of limited altruism. For the dispositional theorist is then in a position to say not just that our moral motivations can be given a privileged rational defence to the extent that they are based on such deliberations, but also that our moral beliefs can be given a privileged rational defence to just the extent that our moral motivations can. For the fact that we can give a privileged rational defence of our moral beliefs, if indeed we can, simply follows from the fact that the desire in terms of which we analyse rightness is one all rational creatures would have if they reasoned in accordance with principles like the principle of limited altruism.<sup>41</sup>

Reconciliation in the form of the dispositional theory of value is thus the preferred solution to the moral problem. For it alone explains why our reasons for our moral beliefs and our reasons for our moral motivations are one and the same. Our moral beliefs and motivations are each justified to the extent that they are based on reasoning in accordance with valid principles that permit us to derive evaluative truths from truths about our circumstances.

Of course, the dispositional theory merely offers us a reconciliation of (1), (2) and (3) at the level of *concepts*, not *ontology*. If no such principles are valid then moral judgements of the form ' $\Phi$ -ing is right' are all, strictly speaking, false. It might

<sup>41</sup> I discuss these matters in my 'Objectivity and Moral Realism'.

be thought that, if this turns out to be so, we would then be forced to adopt an error theory, in the spirit of John Mackie. But that is not our only option. We might instead conclude that rightness isn't everything we thought it was, so opting for *conceptual revision*.

Revisionists who urge rejection of (1) might insist that this is the conclusion they reached long ago. Revisionists who urge rejection of (2) might now insist that their view be thought of as a serious competitor in this quest for revision. I do not myself think that, even if we did opt to revise our beliefs about what rightness is, we would have to respond in either of these ways, however. For we might prefer instead to respond by giving an anti-rationalist relativistic dispositional theory of rightness.<sup>42</sup> But even if we did I want to emphasise how *different* such revisionist strategies look from the way they looked earlier. For it now seems quite inappropriate for those who opt to revise by rejecting (1) to say that they reject (1) rather than (2) because they accept (2) as *basic*, and likewise for those who say that they opt to revise by rejecting (2) because they take the *denial* of (2) to be basic. (1) and (2) are intertwined, the appeal of each lying in a conception of reasons that at one and the same time can be reasons for our moral beliefs and reasons for our moral motivations. Revisionists who wish to cash in on the failure of this form of the dispositional theory, if indeed it fails, thus have to admit that we *should*, but *can't*, accept *both* (1) and (2). Reconciliation must be given its due even if we ultimately opt for revision. Once this is agreed I am happy enough to let the revisionists fight it out amongst themselves for final honours as regards the moral problem.<sup>43</sup>

<sup>42</sup> See my 'Should We Believe in Emotivism?' pp. 302-5 and footnote 22.

<sup>43</sup> Thanks to Rory and Owen Pettit for footnote 15 and Eileen McNally for her help in transcription. This paper was largely written while I was a Visiting Fellow in the Department of Philosophy, Research School of Social Sciences, Australian National University. I thank them for their support.

## DISPOSITIONAL THEORIES OF VALUE

Michael Smith, David Lewis and Mark Johnston

*II—David Lewis*

Roughly, values are what we are disposed to value. Less roughly, we have this schematic definition: *Something of the appropriate category is a value if and only if we would be disposed, under ideal conditions, to value it.* It raises five questions. (1) What is the favourable attitude of ‘valuing’? (2) What is the ‘appropriate category’ of things? (3) What conditions are ‘ideal’ for valuing? (4) Who are ‘we’? (5) What is the modal status of the equivalence?

By answering these questions, I shall advance a version of the dispositional theory of value. I begin by classifying the theory that is going to emerge. First, it is naturalistic: it advances an analytic definition of value. It is naturalistic in another sense too: it fits into a naturalistic metaphysics. It invokes only such entities and distinctions as we need to believe in anyway, and needs nothing extra before it can deliver the values. It reduces facts about value to facts about our psychology.

The theory is subjective: it analyses value in terms of our attitudes. But it is not subjective in the narrower sense of implying that value is a topic on which whatever we may think is automatically true, or on which there is no truth at all. Nor does it imply that if we had been differently disposed, different things would have been values. Not quite—but it comes too close for comfort.

The theory is internalist: it makes a conceptual connection between value and motivation. But it offers no guarantee that everyone must be motivated to pursue whatever is of value; still less, whatever he judges to be of value. The connection is defeasible, in more ways than one.

The theory is cognitive: it allows us to seek and to gain knowledge about what is valuable. This knowledge is a *posteriori* knowledge of contingent matters of fact. It could in principle be gained by psychological experimentation. But it is more likely to be gained by difficult exercises of imagination, carried out perhaps in a philosopher’s or a novelist’s armchair.



The theory is conditionally relativist: it does not exclude the possibility that there may be no such thing as value *simpliciter*, just value for this or that population. But it does not imply relativity, not even when taken together with what we know about the diversity of what people actually value. It leaves the question open.

Is it a form of realism about value?—That question is hard. I leave it for the end.

*What is 'valuing'?* It is some sort of mental state, directed toward that which is valued. It might be a feeling, or a belief, or a desire. (Or a combination of these; or something that is two or three of them at once; or some fourth thing. But let us set these hypotheses aside, and hope to get by with something simpler.<sup>1</sup>)

A feeling?—Evidently not, because the feelings we have when we value things are too diverse.

A belief? What belief? You might say that one values something just by believing it to be a value. That is circular. We might hide the circularity by maneuvering between near-synonyms, but it is better to face it at once. If so, we have that being a value is some property such that something has it iff we are disposed, under ideal conditions, to believe that the thing has it. In other words, such that we are disposed, under ideal

<sup>1</sup> The most interesting of the hypotheses here set aside is that an attitude of valuing might be a 'besire': a special kind of attitude that is both a belief and a desire and that motivates us, without benefit of other desires, in just the way that ordinary desires do. (Or it might be an attitude that is not identical with, but rather is necessarily connected with, a belief and a desire; or an attitude that is not strictly speaking either a belief or a desire, but is just like each apart from also being like the other.) Valuing X might be the besire that is at once a belief that X is good and a desire for X; where *goodness* just means that property, whatever it may be, such that a belief that X has it may double as a desire for X.

But we should hesitate to believe in besires, because integrating them into the folk psychology of belief and desire turns out to be no easy thing. On the difficulty with instrumental besires, see my 'Desire as Belief' and John Collins, 'Belief, Desire and Revision', *Mind* 97 (1988), pp. 323–342: when a system of attitudes changes under the impact of new information, beliefs evolve in one way and (instrumental) desires in another. A besire, trying to go both ways at once, would be torn apart. Intrinsic besires—a better candidate for the attitude of valuing—face a different difficulty. At least in miniature examples, they turn out to be altogether impervious to change under the impact of experience. Not bad, you might think—why *should* experience change our mind about what's intrinsically good? The trouble is that the result applies not only to perceptual experience but also to experience of moral reflection, 'intuiting', and the like.

conditions, to be right about whether something has it. That is not empty; but it tells us little, since doubtless there are many properties about which we are disposed to be right.

Further, if valuing something just meant having a certain belief about it, then it seems that there would be no conceptual reason why valuing is a *favourable* attitude. We might not have favoured the things we value. We might have opposed them, or been entirely indifferent.

So we turn to desires. But we'd better not say that valuing something is just the same as desiring it.<sup>2</sup> That may do for some of us: those who manage, by strength of will or by good luck, to desire exactly as they desire to desire. But not all of us are so fortunate. The thoughtful addict may desire his euphoric daze, but not value it. Even apart from all the costs and risks, he may hate himself for desiring something he values not at all. It is a desire he wants very much to be rid of.<sup>3</sup> He desires his high, but he does not desire to desire it, and in fact he desires not to desire it. He does not desire an unaltered, mundane state of consciousness, but he does desire to desire it. We conclude that he does not value what he desires, but rather he values what he desires to desire.

Can we do better by climbing the ladder to desires of ever-higher order? What someone desires to desire to desire might conceivably differ from what he does desire to desire. Or. . . . Should we perhaps say that what a person really values is given by his highest order of desire, whatever order that is?—It is hard to tell whether this would really be better, because it is hard to imagine proper test cases.<sup>4</sup> Further, if we go for the highest

<sup>2</sup> Often in decision theory and economics, 'value' does just mean a measure of desiredness, and all desires count equally. But it's not the sense we want here.

<sup>3</sup> On desires to desire, see Harry Frankfurt, 'Freedom of the Will and the Concept of a Person', *Journal of Philosophy* 68 (1971), pp. 5-20; and Richard C. Jeffrey, 'Preference Among Preferences', *Journal of Philosophy* 71 (1974), pp. 377-391.

<sup>4</sup> It is comparatively easy to imagine *instrumental* third-order desires. Maybe our addict wishes he could like himself better than he does; and not by doing away with his addiction, which he takes to be impossible, but by becoming reconciled to it and accepting himself as he is. Or maybe he just fears that his second-order desire not to be addicted will someday lead him to suffer the pains of withdrawal. Either way, he wants to be rid of his second-order desire not to be addicted, but he wants it not for itself but as a means to some end. This is irrelevant: presumably it is intrinsic, not instrumental, desiring that is relevant to what someone values.

order, we automatically rule out the case of someone who desires to *value* differently than he does, yet this case is not obviously impossible. I hesitantly conclude we do better to stop on the second rung: valuing is just desiring to desire.

Recall G. E. Moore: ‘To take, for instance, one of the more plausible, because one of the more complicated, of such proposed definitions, it may easily be thought, at first sight, that to be good may mean to be that which we desire to desire’.<sup>5</sup> Of course he does not endorse the definition, but at least he does it the honour of choosing it for his target to display the open question argument. I don’t say that everything we value is good; but I do echo Moore to this extent. I say that to be *valued* by us means to be that which we desire to desire. Then to be a value—to be good, near enough—means to be that which we are disposed, under ideal conditions, to desire to desire. Still more complicated, still more plausible. It allows, as it should, that under less-than-ideal conditions we may wrongly value what is not really good. As for Moore’s open question, we shall face that later.

We have this much of an ‘internalist’ conceptual connection between value and motivation. If something is a value, and if someone is one of the appropriate ‘we’, and if he is in ideal conditions, then it follows that he will value it. And if he values it, and if he desires as he desires to desire, then he will desire it. And if he desires it, and if this desire is not outweighed by other conflicting desires, and if he has the instrumental rationality to do what serves his desires according to his beliefs, then he will pursue it. And if the relevant beliefs are near enough true, then he will pursue it as effectively as possible. A conceptual connection between value and motivation, sure enough—but a multifariously iffy connection. Nothing less iffy would be credible. But still less is it credible that there is no connection at all.

In general, to find out whether something is disposed to give response R under conditions C, you can put it in C and find out whether you get R. That is a canonical way to learn whether the disposition is present, though surely not the only possible

<sup>5</sup> *Principia Ethica* (Cambridge University Press, 1903) Section 13.

way.<sup>6</sup> If a dispositional theory of value is true, then we have a canonical way to find out whether something is a value. To find out whether we would be disposed, under ideal conditions, to value it, put yourself in ideal conditions, if you can, making sure you can tell when you have succeeded in doing so. Then find out whether you value the thing in question, i.e. whether you desire to desire it. If you do, that confirms that it is a value. (I assume you are one of the appropriate 'we' and you know it.) Now we have this much of an 'internalist' conceptual connection between value judgements and motivation. It is even iffier than the connection between value itself and motivation; and again I say that if it were less iffy, it would be less credible. If someone believes that something is a value, and if he has come to this belief by the canonical method, and if he has remained in ideal conditions afterward or else retained the desire to desire that he had when in ideal conditions, then it follows that he values that thing. And if he desires as he desires to desire, then he desires that thing; and so on as before.

The connection is not with the judgement of value *per se*, but with the canonical way of coming to it. If someone reached the same judgement in some non-canonical way—as he might—that would imply nothing about his valuing or desiring or pursuing.

*What is the 'appropriate category'?* If values are what we are disposed to desire to desire, then the things that can be values must be among the things that can be desired. Those fall into two classes.

Sometimes, what one desires is that the world should be a

<sup>6</sup> It is a fallible way; for it may be that you cannot put the thing in C without making the disposition disappear. Imagine that a surface now has just the molecular structure that disposes things to reflect light; but that exposing it to light would catalyze a swift chemical change and turn it into something unreflective. So long as it's kept in the dark, is it reflective?—I think so; but its reflectivity is what Ian Hunt once called a 'finkish' disposition, one that would vanish if put to the test. (So a simple counterfactual analysis of dispositions fails.) Could a disposition to value, or to disvalue, be finkish? Yes; here is an example due to Michael Tooley. Suppose, as I shall claim, that 'ideal conditions' include imaginative acquaintance; suppose there is no way to imagine direct electrical stimulation of the pleasure centre of the brain except by trying it out; and suppose that one brief trial would enslave you to the electrode and erase all other desires. Then I think you might well have a finkish disposition to disvalue the experience. If, *per impossibile*, you could manage to imagine it without at the same time having your present system of desires erased by the current, you would desire not to desire it.

certain way: that it should realise one of a certain class of (maximally specific, qualitatively delineated) possibilities for the whole world. This class—a ‘proposition’, in one sense of that word—gives the content of the desire. To desire that the world realise some possibility within the class is to desire that the proposition be true. Call this ‘desire *de dicto*’.

But sometimes, what one desires concerns not just the world but oneself: one simply desires to *be* a certain way. For instance, Fred might want to be healthy, or wealthy, or wise. Then what he wants is that he himself should realise one of a certain class of (maximally specific, qualitatively delineated) possibilities for an individual—or better, for an individual-in-a-world-at-a-time. This class—a ‘property’ in one sense of that word, or an ‘egocentric proposition’—gives the content of the desire. To desire to realise some possibility in the class is to desire to have the property, or to desire that the egocentric proposition be true of one. Call this ‘desire *de se*’, or ‘egocentric’ or ‘essentially indexical’ desire.<sup>7</sup>

You might think to reduce desire *de se* to desire *de dicto*, saying that if Arthur desires to be happy, what he desires is that the world be such that Arthur is happy. (You might doubt that such worlds comprise a qualitatively delineated class, so you might consider dropping that requirement.) But no. That is not exactly the same thing, though the difference shows up only when we imagine someone who is wrong or unsure about who in the world he is. Suppose Arthur thinks he is Martha. If Arthur is self-centred he may desire to be happy, desire that the world be one wherein Martha is happy, but not desire that the world is one wherein Arthur is happy. If instead Arthur is selflessly benevolent he may not desire to be happy, yet he may desire that the world be such that Arthur is happy. If Arthur is so befuddled as not to know whether he is Arthur or Martha, but hopes he is Arthur, he does not just desire that the world be such that

<sup>7</sup> See Peter Geach, ‘On Beliefs about Oneself’, *Analysis* 18 (1957), pp. 23–24; Hector-Neri Castañeda ‘On the Logic of Attributions of Self-Knowledge to Others’, *Journal of Philosophy* 65 (1968), pp. 439–456; John Perry, ‘Frege on Demonstratives’, *Philosophical Review* 86 (1977), pp. 474–497, and ‘The Problem of the Essential Indexical’, *Noûs* 14 (1979), pp. 3–21; my ‘Attitudes *De Dicto* and *De Se*’, *Philosophical Review* 88 (1979), pp. 513–543; Roderick Chisholm, *The First Person: An Essay on Reference and Intentionality* (Harvester Press, 1981).

Arthur is self-identical! In all these cases, Arthur's desire is, at least in part, irreducibly *de se*.<sup>8,9</sup>

When we acknowledge desires *de se*, we must distinguish two senses of 'desiring the same thing'. If Jack Sprat and his wife both prefer fat meat, they *desire alike*. They are psychological duplicates, on this matter at least. But they do not *agree* in their desires, because no possible arrangement could satisfy them both. Whereas if Jack prefers the fat and his wife prefers the lean, then they differ psychologically, they do not desire alike. But they do agree, because if he eats no fat and she eats no lean, that would satisfy them both. In general, they desire alike iff they desire *de se* to have exactly the same properties and they desire *de dicto* that exactly the same propositions hold. They agree in desires iff exactly the same world would satisfy the desires of both; and a world that satisfies someone's desires is one wherein he has all the properties that he desires *de se* and wherein all the propositions hold that he desires *de dicto*. Agreement in desire makes for harmony; desiring alike may well make for strife.

As we can desire *de dicto* or *de se*, so we can desire to desire *de dicto* or *de se*. If desiring to desire is valuing, and if values are what we are disposed to value, then we must distinguish values *de dicto* and *de se*. A value *de dicto* is a proposition such that we are disposed to desire to desire *de dicto* that it hold. A value *de se* is a property such that we are disposed to desire to desire *de se* to have it.

It is essential to distinguish. Consider egoism: roughly, the thesis that one's own happiness is the only value. Egoism is meant to be general. It is not the thesis that the happiness of a certain special person, say Thrasymachus, is the only value. Egoism *de dicto* says that for each person X, the proposition that

<sup>8</sup> What we can do is to go the other way, subsuming desire *de dicto* under desire *de se*. To desire that the world be a certain way is to desire that one have the property of living in a world that is that way—a property that belongs to all or none of the inhabitants of the world, depending on the way the world is. This subsumption, artificial though it be, is legitimate given a suitably broad notion of property. But for present purposes we need distinction, not unification. So let us henceforth ignore those desires *de se* that are equivalent to desires *de dicto*, and reserve the term '*de se*' for those that are not.

<sup>9</sup> If you like, you can put the egocentricity not in the content of desire itself but in an egocentric mode of presentation of that content. The choice matters little, save to simplicity. See Jeremy Butterfield, 'Content and Context' in Butterfield, ed., *Language Mind and Logic* (Cambridge University Press, 1986).

X is happy is the only value. That is inconsistent, as Moore observed.<sup>10</sup> It says that there are as many different values as there are people, and each of them is the only value. Egoism *de se* says that the property of happiness—in other words, the egocentric proposition that one is happy—is the only value. Moore did not confute that. He ignored it. False and ugly though it be, egoism *de se* is at least a consistent doctrine. What it alleges to be the only value would indeed be just *one* value *de se*, not a multitude of values *de dicto*.<sup>11</sup>

Insofar as values are *de se*, the wholehearted pursuit by everyone of the same genuine value will not necessarily result in harmony. All might value alike, valuing *de se* the same properties and valuing *de dicto* the same propositions. Insofar as they succeed in desiring as they desire to desire, they will desire alike. But that does not ensure that they will agree in desire. If egoism *de se* were true, and if happiness could best be pursued by doing others down and winning extra shares, then the pursuit by all of the very same single value would be the war of all against all.

Because egoism is false and ugly, we might be glad of a theoretical framework that allowed us to confute it *a priori*. And some of us might welcome a framework that promises us harmony, if only we can all manage to pursue the same genuine values. Was it right, then, to make a place for values *de se*? Should we have stipulated, instead, that something we are disposed to desire to desire shall count as a value only when it is a proposition that we are disposed to desire to desire *de dicto*?

No. Probably it is already wrong to reject egoism *a priori* but, be that as it may, there are other doctrines of value *de se*, more plausible and more attractive. Self-improvement and self-sacrifice are no less egocentric than self-aggrandizement and

<sup>10</sup> *Principia Ethica*, Section 59.

<sup>11</sup> Someone who said that happiness was the only value might mean something else, which is not a form of egoism at all. He might mean that the proposition that happiness is maximized is the only value—a single value *de dicto*. Or he might mean that for each person X, the proposition that X is happy is a value *de dicto*, and that these many values of parallel form are the only values. Mean what you please—I take these to be legitimate, but derivative, senses in which a property may be called a value. I only say they should not be confused with, or drive out, the sense in which a property may be a value *de se*.

self-indulgence. Surely we should make a place for putative values *de se* of altruism, of honour, and of loyalty to family, friends, and country.<sup>12</sup> We may entertain the substantive thesis that none of these putative values *de se* is genuine, and that all genuine values are *de dicto*. But even if we believed this—myself, I think it wildly unlikely—we should not beg the question in its favour by building it into our theoretical framework.

*What conditions are 'ideal'?* If someone has little notion what it would be like to live as a free spirit unbound by law, custom, loyalty, or love; or what a world of complete harmony and constant agreement would be like; then whether or not he blindly values these things must have little to do with whether or not they are truly values. What he lacks is imaginative acquaintance. If only he would think harder, and imagine vividly and thoroughly how it would be if these putative values were realised (and perhaps also how it would be if they were not) that would make his valuing a more reliable indicator of genuine value. And if he could gain the fullest imaginative acquaintance that is humanly possible,<sup>13</sup> then, I suggest, his valuing would be an infallible indicator. Something is a value iff we are disposed, under conditions of the fullest possible imaginative acquaintance, to value it.

Compare a version of Intuitionism: by hard thought, one becomes imaginatively well acquainted with X; in consequence, but not as the conclusion of any sort of inference, one intuits that X has a certain unanalysable, non-natural property; and in consequence of that, one comes to value X. My story begins and ends the same. Only the middle is missing. Again, an exercise of imaginative reason plays a crucial role. Again, its relation to what follows is causal, and in no way inferential. But in my story, the consequent valuing is caused more directly, not via the detection of a peculiar property of X.

Can we say that the valuing ensued because X was a value?—Maybe so, but if we do, we are not saying much: it

<sup>12</sup> See Andrew Oldenquist, 'Loyalties', *Journal of Philosophy* 79 (1982), pp. 173–193; Michael Slote, 'Morality and Self-Other Asymmetry', *Journal of Philosophy* 81 (1984), pp. 179–192.

<sup>13</sup> Without in the process having his dispositions to value altered—see Footnote 6.



ensues because there is something about imaginative acquaintance with X that causes valuing.<sup>14</sup>

The canonical way to find out whether something is a value requires a difficult imaginative exercise. And if you are to be sure of your answer, you need to be sure that you have gained the fullest imaginative acquaintance that is humanly possible. A tall order! You had better settle for less. Approximate the canonical test. Try hard to imagine how it would be if the putative value were (or were not) realised. Hope that your acquaintance comes close enough to the fullest possible that getting closer would not change your response. Then you may take your valuing as fallible evidence that you were acquainted with a genuine value, or your indifference as fallible evidence that you were not. You cannot be perfectly certain of your answer, but you can take it as sure enough to be going on with, subject to reconsideration in the light of new evidence. How sure is that?—Well, as always when we acknowledge fallibility, some of us will be bolder than others.

New evidence might be a more adequate imaginative exercise of your own. It might be the testimony of others. It might in principle be a result of scientific psychology—though it is far from likely that any such results will come to hand soon!

A trajectory toward fuller imaginative acquaintance with putative value X is not just a sequence of changes in your imaginative state. It has a direction to it. And that is so independently of my claim that it leads, after a point, to ever-surer knowledge about whether X is a value. For in learning how to imagine X, you gain abilities; later you have all the relevant imaginative abilities you had before, and more besides. And you notice, *a priori*, relationships of coherence or incoherence between attitudes that might figure in the realisation of X; later you are aware of all that you had noticed before, and more

<sup>14</sup>How does imaginative acquaintance cause valuing, when it does? How does imagination render values attractive? Does it happen the same way for all values?—For our purposes, it is enough to say that it happens. We needn't know how. But we may guess. Maybe imaginative acquaintance shows us how new desires would be seamless extensions of desires we have already. Or maybe we gravitate toward what we understand, lest we baffle ourselves—see J. David Velleman, *Practical Reflection* (Princeton University Press, forthcoming). But that cannot be the whole story, because some easily understood lives—say a life of lethargy, ruled by a principle of least action—remain repellent.

besides. And you think of new questions to explore in your imagining—what might the life of the free spirit become, long years after its novelty had worn off?—and later you have in mind all the questions you had thought of before, and more besides. Forgetting is possible, of course. But by and large, the process resists reversal.<sup>15</sup>

Our theory makes a place for truth, and in principle for certain knowledge, and in practice for less-than-certain knowledge, about value. But also it makes a place for ignorance and error, for hesitant opinion and modesty, for trying to learn more and hoping to succeed. That is all to the good. One fault of some subjective and prescriptive theories is that they leave no room for modesty: just decide where you stand, then you may judge of value with the utmost confidence!

There is a long history of theories that analyse value in terms of hypothetical response under ideal conditions, with various suggestions about what conditions are ideal. Imaginative acquaintance often gets a mention. But much else does too. I think imaginative acquaintance is all we need—the rest should be in part subsumed, in part rejected.

First, the responder is often called an ideal *spectator*. That is tantamount to saying that conditions are ideal only when he is observing a sample of the putative value in question (or of its absence). If the putative value is *de se*, a property, then a sample can just be an instance. If it is *de dicto*, a proposition, it is hard to say in general what an observable sample could be. But if it is the proposition that a certain property is instantiated sometimes, or often, or as often as possible, or in all cases of a certain kind, then again a sample can just be an instance of the property. Anyone happy may serve as a sample of the proposition that total happiness is maximised.

Observable samples can sometimes prompt the imagination and thereby help us to advance imaginative acquaintance. But they are of limited use. For one thing, observation does not include mind-reading. Also, it does best with short, dramatic episodes. A lifelong pattern of stagnation, exemplifying the absence of various values, goes on too long to be easily

<sup>15</sup> For a discussion of unidirectionality in aesthetic valuing, see Michael Slote, 'The Rationality of Aesthetic Value Judgements', *Journal of Philosophy* 68 (1971), pp. 821–839.

observed. Samples are dispensable as aids to imagination, and sometimes they are comparatively ineffective. A novel might be better.

The notion of an ideal spectator is part of a longstanding attempt to make dispositional theories of value and of colour run in parallel. But the analogy is none too good, and I doubt that it improves our understanding either of colour or of value. Drop it, and I think we have no further reason to say that a disposition to value is a disposition to respond to observed samples.<sup>16</sup>

Second, the ideal responder is often supposed to be well informed. If any item of empirical knowledge would affect his response, he knows it.—But some sorts of knowledge would not help to make your valuing a more reliable indicator of genuine value. Instead they would distract. If you knew too well how costly or how difficult it was to pursue some value, you might reject the grapes as sour, even when imaginative acquaintance with the value itself would have caused you to value it. Genuine values might be unattainable, or unattainable without undue sacrifice of other values. An ideal *balancer* of values needs thorough knowledge of the terms of trade. An ideal valuer may be better off without it. Our present business is not with the balancing, but with the prior question of what values there are to balance.<sup>17</sup>

Another unhelpful sort of knowledge is a vivid awareness that we are small and the cosmos is large; or a vivid awareness of the mortality of mankind, and of the cosmos itself. If such knowledge tends to extinguish all desire, and therefore all

<sup>16</sup> If we had demanded samples, we would have had a choice about where to locate the disposition. Is it within us or without? Is it a disposition in the samples to evoke a response from spectators?—that is what best fits the supposed parallel with a dispositional theory of colour. See Robert Pargetter and John Campbell, 'Goodness and Fragility', *American Philosophical Quarterly* 23 (1986), pp. 155–166, for an analysis of this kind. Or is it a disposition in the spectators to respond to samples? Or is it a disposition of the sample-cum-spectator system to respond to having its parts brought together? For us there is no choice. The propositions and properties that are the values cannot harbour any causal bases for dispositions. Samples could, but there needn't be any samples. Imaginative experiences could, but those are within us, and are not themselves samples of values. So the disposition must reside in us, the responders. Being a value comes out as a dispositionally analysed property, but not as a disposition of the things that have it. Values themselves are not disposed to do anything.

<sup>17</sup> Previous theories of hypothetical response may indeed have been concerned as much with the analysis of right balancing as with value itself. If so, they cannot be faulted for trying to characterise an ideal balancer. However my present *analysandum* is different.

valuing, it will not help us to value just what is valuable. Likewise it will be unhelpful to dwell too much on the lowly causal origins of things. If some feature of our lives originated by kin selection, or Pavlovian conditioning, or sublimation of infantile sexuality, that is irrelevant to what it is like in itself. Unless he can overcome the illusion of relevance, a valuer will be more reliable if he remains ignorant of such matters.

However, I grant one case—a common one—in which one does need empirical knowledge in order to gain imaginative acquaintance with a given putative value. It may be ‘given’ in a way that underspecifies it, with the rest of the specification left to be filled in by reference to the actual ways of the world. For instance when I mentioned the life of a free spirit as a putative value, what I meant—and what you surely took me to mean—was the life of a free spirit in a world like ours. In such cases, a valuer must complete the specification by drawing on his knowledge of the world, else he will not know what he is supposed to imagine. To that extent—and only to that extent, I think—being well-informed is indeed a qualification for his job.<sup>18</sup>

Third, it may be said that the ideal responder should not only imagine having (or lacking) a putative value, but also imagine the effect on other people of someone’s having (or lacking) it. Thinking what it would be like to live as a free spirit is not enough. You must also think what it would be like to encounter the free spirit and be ill-used.—But again, I think the requirement is misplaced. It is appropriate not to an ideal valuer, but to an ideal balancer who must think through the cost to some values of the realisation of others. In addressing the prior question of what values there are, counting the cost is a distraction to be resisted.

Often, however, realising a putative value *de se* would itself involve imagining the impact of one’s conduct on other people. When that is so, imagining realising the value involves

<sup>18</sup> Imaginative acquaintance is sometimes thought to consist in the possession of a special kind of ‘phenomenal’ information. If that is so, of course my own candidate for ‘ideal conditions’ comes down to a special case of being well-informed. But it is not so—not even in the most favourable case, that of imaginative acquaintance with a kind of sense-experience. See my ‘What Experience Teaches’ in William Lycan, ed., *Mind and Cognition: A Reader* (Blackwell, 1989).

imagining imagining the impact; and that cannot be done without simply imagining the impact. In such cases, imagining the impact does fit in; for it is already subsumed as part of imaginative acquaintance with the value itself.

Fourth, the ideal responder is often said to be dispassionate and impartial, like a good judge.—Once more, the requirement is appropriate not to an ideal valuer but to an ideal balancer. The valuer is not a judge. He is more like an advocate under the adversarial system. He is a specialist, passionate and partial perhaps, in some one of all the values there are. On the present theory, when I say that X is a value iff we are disposed to value X under ideal conditions, I do not mean conditions that are ideal *simpliciter*, but rather conditions that are ideal *for* X. We should not assume that there is any such thing as a condition of imaginative acquaintance with all values at once. (Still less, all putative values.) Imagination involves simulation—getting into the skin of the part. How many skins can you get into all at once? Tranquillity and vigorous activity might both be values; but a full imaginative acquaintance with one might preclude a full imaginative acquaintance with the other. (The incompatibility might even be conceptual, not just psychological.) Then if we value both, as surely many of us do, it is not because of acquaintance with both at once. It might be a lasting effect of past imaginative acquaintance at some times with one and at other times with the other.

A further speculation: it might happen that there were values that could not even be valued all at once. If so, then conflict of values would go deeper than is ever seen in hard choices; because what makes a choice hard is that conflicting values *are* valued together by the unfortunate chooser. An alarming prospect!—or exhilarating, to those of us who delight in the rich variety of life.

*Who are 'we'?* An *absolute* version of the dispositional theory says that the 'we' refers to all mankind. To call something a value is to call it a value *simpliciter*, which means that everyone, always and everywhere, is disposed under ideal conditions to value it. Then there are values only insofar as all mankind are alike in their dispositions.

Maybe all mankind *are* alike. The manifest diversity of

valuing between different cultures—or for that matter within a culture, say between colleagues in the same philosophy department—is no counterevidence. In the first place, people may not be valuing as they would be disposed to value under ideal conditions. In the second place, remember that conditions of imaginative acquaintance are ideal for particular values, not *simpliciter*. So even if all are disposed alike, and all value as they would under ideal conditions, that may mean that some people value X as they would under conditions ideal for X, while others, who are no differently disposed, value Y as they would under conditions ideal for Y. If no conditions are ideal at once for X and for Y (still more if X and Y cannot both be valued at once), there could be diversity of valuing even in a population of psychological clones, if different ones had been led into different imaginative exercises.

We saw that it would be no easy job to find out for sure whether a particular person would be disposed to value something under ideal conditions of imaginative acquaintance with it. It would be harder still to find out all about one person's dispositions. And not just because one hard job would have to be done many times over. It might happen that imaginative acquaintance with X would leave traces, in one's valuing or otherwise, that got in the way of afterward imagining Y. To the extent that there was such interference, each new imaginative experiment would be harder than the ones before.

The fallback, if we are wary of presupposing that all mankind are alike in their dispositions to value, is tacit relativity. A *relative* version says that the 'we' in the analysis is indexical, and refers to a population consisting of the speaker and those somehow like him. If the analysis is indexical, so is the analysandum. Then for speaker S to call something a value is to call it a value for the population of S and those like him; which means that S and those like him are all disposed, under ideal conditions, to value it.

The relative version is not just one version, but a spectrum. What analysis you get depends on how stringent a standard of similarity you apply to the phrase 'the speaker and those somehow like him'. At one end of the spectrum stands the absolute version: common humanity is likeness enough, so whoever speaks, all mankind are 'we'. At the other end, 'we'

means: 'you and I, and I'm none too sure about you'. (Or it might be 'I, and those who think as I do', which reduces to 'I'.) In between, 'we' means: 'I, and all those who are of a common culture with me'. Since mankind even at one moment is not made up of isolated and homogeneous tribes, and since we should not limit ourselves to the part of mankind located at one moment, we may haggle endlessly over how much cultural affiliation is meant.

(We have a piece of unfinished business: if someone is to find out about values by the canonical method, he must somehow know that he is one of the appropriate 'we'. All our versions, absolute or relative, make this knowledge automatic. Not so for *elitist* versions, on which 'we' means 'the best-qualified of us' or maybe 'the most normal of us'. But elitist versions are pointless. We're already considering dispositions under extravagantly ideal conditions; we needn't idealise all over again by being selective about who counts as one of the 'we'.)

If some relative version were the correct analysis, wouldn't that be manifest whenever people talk about value? Wouldn't you hear them saying 'value for me and my mates' or 'value for the likes of you'? Wouldn't you think they'd stop arguing after one speaker says X is a value and the other says it isn't?—Not necessarily. They might always presuppose, with more or less confidence (well-founded or otherwise), that whatever relativity there is won't matter in *this* conversation. Even if they accept in principle that people sometimes just differ in their dispositions to value, they may be very reluctant to think the present deadlocked conversation is a case of such difference. However intractable the disagreement may be, they may go on thinking it really *is* a disagreement: a case in which two people are disposed alike, but one of them is wrong about what is a value relative to their shared dispositions, because he is not valuing as he would under ideal conditions. So long as they think that—and they might think it very persistently—they can hold the language of explicit relativity in reserve. It is there as a last resort, if ever they meet with a proven case of ultimate difference. But it will not be much heard, since it is a practical impossibility to prove a case. If the language of absolutism prevails, that is not strong evidence against relativity.

(Those who have heard of the relativity of simultaneity do not

manifest this knowledge all the time. They speak as the ignorant do, and no harm done. They'll resort to the language of relativity when it matters, say in discussing the exploits of the interstellar navy.)

*Does* the language of absolutism prevail? Not really. With some of us it does. Others of us resort to the language of relativity at the drop of a hat. Yet this too is poor evidence. The eager relativists may have been confused by philosophy. For who can escape it?

So what version should we prefer, absolute or relative?—Neither; instead, I commend a *wait-and-see* version. In making a judgement of value, one makes many claims at once, some stronger than others, some less confidently than others, and waits to see which can be made to stick. I say X is a value; I mean that all mankind are disposed to value X; or anyway all nowadays are; or anyway all nowadays are except maybe some peculiar people on distant islands; or anyway . . . ; or anyway you and I, talking here and now, are; or anyway I am.<sup>19</sup> How much am I claiming?—as much as I can get away with. If my stronger claims were proven false—though how that could be proven is hard to guess—I still mean to stand by the weaker ones. So long as I'm not challenged, there's no need to back down in advance; and there's no need to decide how far I'd back down if pressed. What I mean to commit myself to is *conditionally relative*: relative if need be, but absolute otherwise.

*What is the modal status of the equivalence?* The equivalence between value and what we are disposed to value is meant to be a piece of philosophical analysis, therefore analytic. But of course it is not obviously analytic; it is not even obviously true.

It is a philosophical problem how there can ever be unobvious analyticity. We need not solve that problem; suffice it to say that it is everybody's problem, and it is not to be solved by denying the phenomenon. There are perfectly clear examples of it: the epsilon-delta analysis of an instantaneous rate of change, for one. Whenever it is analytic that all A's are B's, but not obviously analytic, the Moorean open question—whether all A's are indeed B's—is intelligible. And not only is it intelligible

<sup>19</sup> See the discussion of 'anyway' in Frank Jackson, 'On Assertion and Indicative Conditionals', *Philosophical Review* 88 (1979), pp. 565–589.



in the sense that we can parse and interpret it (that much is true even of the question whether all A's are A's) but also in the sense that it makes sense as something to say in a serious discussion, as an expression of genuine doubt.

Besides unobvious analyticity, there is equivocal analyticity. Something may be analytic under one disambiguation but not another, or under one precisification but not another. Examples abound. Quine was wrong that analyticity was unintelligible, right to doubt that we have many clearcut cases of it. If differing versions of a concept (or, if you like, different but very similar concepts) are in circulation under the same name, we will get equivocal analyticity. It is analytic under one disambiguation of 'dog' that all dogs are male; under one disambiguation of 'bitch' that all bitches are canine. It is analytic under some precisifications of 'mountain' that no mountain is less than one kilometre high. When analyticity is equivocal, open questions make good conversational sense: they are invitations to proceed under a disambiguation or precisification that makes the answer to the question not be analytic. By asking whether there are mountains less than one kilometre high, you invite your conversational partners to join you in considering the question under a precisification of 'mountain' broad enough to make it interesting; yet it was analytic under another precisification that the answer was 'no'.<sup>20</sup> So even if all is obvious, open questions show at worst that the alleged analyticity is equivocal.

I suggest that the dispositional theory of value, in the version I have put forward, is equivocally as well as unobviously analytic. I do not claim to have captured the one precise sense that the word 'value' bears in the pure speech, uncorrupted by philosophy, that is heard on the Clapham omnibus. So far as this matter goes, I doubt that speakers untouched by philosophy are found in Clapham or anywhere else. And if they were, I doubt if they'd have made up their minds exactly what to mean any more than the rest of us have. I take it, rather, that the word 'value', like many others, exhibits both semantic variation and semantic indecision. The best I can hope for is that my dispositional theory lands somewhere near the middle of the

<sup>20</sup> See my 'Scorekeeping in a Language Game', *Journal of Philosophical Logic* 8 (1979), pp. 339-359.

range of variation and indecision—and also gives something that I, and many more besides, could be content to adopt as our official definition of the word ‘value’, in the unlikely event that we needed an official definition.

I’ve left some questions less than conclusively settled: the matter of absolute versus relative versus wait-and-see versions, the details of ‘ideal conditions’, the question of admitting values *de se*, the definition of valuing as second-order versus highest-order intrinsic desiring. It would not surprise or disturb me to think that my answers to those questions are only equivocally analytic—but somewhere fairly central within the range of variation and indecision—and that the same could be said of rival answers. Even if no version of the dispositional theory is unequivocally analytic, still it’s fair to hope that some not-too-miscellaneous disjunction of versions comes out analytic under most reasonable resolutions of indeterminacy (under some reasonable precisification of ‘most’ and ‘reasonable’.)

If the dispositional theory is only unobviously and equivocally analytic, why think that it’s analytic at all?—Because that hypothesis fits our practice. (The practice of many of us, much of the time.) It does seem that if we try to find out whether something is a genuine value, we do try to follow—or rather, approximate—the canonical method. We gain the best imaginative acquaintance we can, and see if we then desire to desire it. In investigating values by the canonical method, we ignore any alleged possibility that values differ from what we’re disposed to value. The dispositional theory explains nicely why we ignore it: no such possibility exists.

Now this should sound an alarm. Phenomenalism, behaviourism, and the like might be supported in exactly the same way: we ignore the possibility that our method of investigation deceives us radically, and the alleged explanation is that no such possibility exists. But in those cases, we know better. We know how systematic hallucination might deceive its victim about the world around him, and how a clever actor might deceive everyone he meets about his inner life (and, in both cases, how it might be that experience or behaviour would remain deceptive throughout the appropriate range of counterfactual suppositions). And it doesn’t just strike us that such deception is possible *somehow*. Rather, we can imagine just how it might happen. We

can give a story of deception all the detail it takes to make it convincing. So we must confess that our method of gaining knowledge of the outer world and the inner lives does consist in part of ignoring genuine possibilities—possibilities that cannot credibly be denied.

The case of value is different, because the convincing detail cannot be supplied. Yes, you might think that perhaps the genuine values somehow differ from what we are disposed to value, even under ideal conditions. (Charles Pigden has noted that a misanthrope might think it because he thinks mankind is irretrievably depraved.) The conjecture is not unthinkable; the dispositional theory is not *obviously* analytic; counterexamples are not *obviously* impossible. That is not yet much evidence of possibility. Better evidence would be a detailed story of just how it might happen that something—something specific—is after all a value that we are not disposed to value, or a non-value that we are disposed to value. But I have no idea how to flesh out the story. Without ‘corroborative detail’, insistence that there exist such possibilities is ‘bald and unconvincing’. This time, nothing outweighs the niceness of explaining the ignoring by denying the possibilities allegedly ignored.

*But is it realism?* Psychology is contingent. Our dispositions to value things might have been otherwise than they actually are. We might have been disposed, under ideal conditions, to value seasickness and petty sleaze above all else. Does the dispositional theory imply that, had we been thus disposed, those things would have been values? That seems wrong.

No: we can take the reference to our dispositions as rigidified. Even speaking within the scope of a counterfactual supposition, the things that count as values are those that we are *actually* disposed to value, not those we would have valued in the counterfactual situation. No worries—unless seasickness actually *is* a value, it still wouldn’t have been a value even if we’d been disposed to value it.

This is too swift. The trick of rigidifying seems more to hinder the expression of our worry than to make it go away. It can still be expressed as follows. We might have been disposed to value seasickness and petty sleaze, and yet we might have been no different in how we used the word ‘value’. The reference of ‘our

actual dispositions' would have been fixed on different dispositions, of course, but our way of fixing the reference would have been no different. In one good sense—though not the only sense—we would have meant by 'value' just what we actually do. And it would have been true for us to say 'seasickness and petty sleaze are values'.

The contingency of value has not gone away after all; and it may well disturb us. I think it is the only disturbing aspect of the dispositional theory. Conditional relativity may well disturb us too, but that is no separate problem. What comfort would it be if all mankind just *happened* to be disposed alike? Say, because some strange course of cultural evolution happened to be cut short by famine, or because some mutation of the brain never took place? Since our dispositions to value are contingent, they certainly vary when we take *all* of mankind into account, all the inhabitants of all the possible worlds. Given the dispositional theory, trans-world relativity is inevitable. The spectre of relativity within our own world is just a vivid reminder of the contingency of value.

If wishes were horses, how would we choose to ride? What would it take to satisfy us? Maybe this new version of the dispositional theory would suit us better: values are what we're *necessarily* disposed to value. Then no contingent 'value' would deserve the name; and there would be no question of something being a value for some people and not for others, since presumably what's necessary is *a fortiori* uniform (unless different dispositions to value are built into different people's individual essences, an unlikely story).

What kind of necessity should it be? Not mere deontic necessity—values are what we're disposed to value on pain of being at fault, where the fault in question turns out to consist in failing to be disposed to value the genuine values. That dispositional theory is empty. Its near relatives are nearly empty. And it won't help to juggle terms; as it might be, by calling it 'rational necessity' and then classifying the disposition to value genuine values as a department of 'rationality'. Probably not nomological necessity either—small comfort to think that we were disposed to disvalue seasickness only because, luckily, our neurons are not subject to a certain fifth force of nature that would distort their workings in just the wrong way. It had better

be necessity *simpliciter*, so-called 'metaphysical' necessity.

If we amend the dispositional theory by inserting 'necessarily', we can be much more confident that the 'values' it defines would fully deserve the name—if there were any of them. But it is hard to see how there possibly could be. If a value, strictly speaking, must be something we are necessarily disposed to value, and if our dispositions to value are in fact contingent, then, strictly speaking, there are no values. If Mackie is right that a value (his term is 'objective good') would have to be

sought by anyone who was acquainted with it, not because of any contingent fact that this person, or every person, is so constituted that he desires this end, but just because the end has to-be-pursuedness somehow built into it,

then he is also right to call values 'queer' and to repudiate the error of believing in them.<sup>21</sup> (Replacing 'sought' by 'valued' would not change that.) If we amend the dispositional theory, requiring values to be all that we might wish them to be, we bring on the error theory. The fire is worse than the frying pan.

Is it, after all, out of the question that our dispositions to value might be necessary? If the theory of mind I favour is true, then the platitudes of folk psychology do have a certain necessity—albeit conditional necessity—to them.<sup>22</sup> There are states that play the functional roles specified in those platitudes, and it is in virtue of doing so that they deserve their folk-psychological names. It is not necessary that there should be any states in us that deserve such names as 'pain', 'belief', or 'desire'. But it is necessary that if any states do deserve those names, then they conform to the platitudes. Or rather, they conform well enough. Now suppose that some of the platitudes of folk psychology specified exactly what we were disposed, under ideal conditions, to desire to desire. And suppose those platitudes were non-negotiable: if a system of states did not satisfy them, that would settle that those states did not conform well enough to folk

<sup>21</sup> J. L. Mackie, *Ethics: Inventing Right and Wrong* (Penguin, 1977), p. 40. But note that the queerness Mackie has in mind covers more than just the to-know-it-is-to-love-it queerness described in this passage.

<sup>22</sup> See my 'An Argument for the Identity Theory', *Journal of Philosophy* 63 (1966), pp. 17–25; and D. M. Armstrong, *A Materialist Theory of the Mind* (Routledge and Kegan Paul, 1968).

psychology to deserve the mental names it implicitly defines. Then there would be things we were necessarily disposed to value—on condition that we had mental lives at all!

The suggestion is intelligible and interesting, but too good to be true. For one thing, it only spreads the trouble. Instead of losing the risk that nothing deserves the name of value, we gain the added risk that nothing deserves commonplace folk-psychological names. *Pace* the Churchlands, it's not really credible that there might turn out to be no beliefs, no desires, no pains, . . .<sup>23</sup> For another thing, it proves too much. It denies outright that it's possible for someone to differ from others in his dispositions to value. Yet this does seem possible; and we can flesh out the story with plenty of 'corroborative detail'. This cunning and subtle villain once was as others are; he gained excellent imaginative acquaintance with many values, and valued them accordingly. Now he has gone wrong, and cares not a fig for what he once valued; and yet he has forgotten nothing. (He certainly has not stopped having any mental life deserving of the name.) He hates those who are as he once was, and outwits them all the better because of his superb empathetic understanding of what they hold dear. Could it not happen?—not if the present suggestion were true. So the present suggestion is false. Yet it was the only hope, or the only one I know, for explaining how there might be things we are necessarily disposed to value. The dispositions are contingent, then. And, at least in some tacit way, we know it. If the story of the subtle villain strikes you as a possible story, that knowledge thereby reveals itself.

But if we know better, it is odd that we are disturbed—as I think many of us will be—by a dispositional theory of value, unamended, according to which values are contingent. It feels wrong. Why might that be?—Perhaps because a large and memorable part of our discussion of values consists of browbeating and being browbeaten.<sup>24</sup> The rhetoric would fall flat if we kept in mind, all the while, that it is contingent how we are disposed to value. So a theory which acknowledges that contingency

<sup>23</sup> As argued in Frank Jackson and Philip Pettit, 'In Defence of Folk Psychology', forthcoming in *Philosophical Studies*.

<sup>24</sup> See Ian Hinckfuss, *The Moral Society: Its Structure and Effects* (Australian National University Discussion Papers in Environmental Philosophy, 1987).

cannot feel quite right. You might say that it is unfaithful to the distinctive phenomenological character of lived evaluative thought. Yet even if it feels not right, it may still be right, or as near right as we can get. It feels not quite right to remember that your friends are big swarms of little particles—it is inadequate to the phenomenology of friendship—but still they are.

I suggested earlier that my version of the dispositional theory of value might be equivocally analytic. So might the amended version, on which values are what we are necessarily disposed to value. Between these two versions, not to mention others, there might be both semantic variation and semantic indecision. If so, it is part of a familiar pattern. One way to create indeterminacy and equivocal analyticity is to define names implicitly in terms of a theory (folk or scientific), and later find out that the theory is wrong enough that nothing perfectly deserves the names so introduced, but right enough that some things, perhaps several rival candidates, deserve the names imperfectly. Nothing perfectly deserves the name 'simultaneity', since nothing quite fits the whole of our old conception. So the name will have to go to some imperfect deserver of it, or to nothing. What it takes to deserve this name, not perfectly but well enough, was never officially settled. One resolution of the indeterminacy makes it analytic that simultaneity must be frame-independent; another, that it must be an equivalence relation; a third, that it must be both at once. The third brings with it an error theory of simultaneity.<sup>25</sup>

I suggest that (for some of us, or some of us sometimes) the amended dispositional theory best captures what it would take for something to perfectly deserve the name 'value'. There are no perfect deservers of the name to be had. But there are plenty of imperfect deservers of the name, and my original version is meant to capture what it takes to be one of the best of them. (But I do not say mine is the only version that can claim to do so. Doubtless there are more dimensions of semantic variation and indeterminacy than just our degree of tolerance for imperfection.) Strictly speaking, nothing shall get the name without deserving it perfectly. Strictly speaking, Mackie is right: genuine values

<sup>25</sup> See Hartry Field, 'Theory Change and the Indeterminacy of Reference', *Journal of Philosophy* 70 (1973), pp. 462-481.

would have to meet an impossible condition, so it is an error to think there are any. Loosely speaking, the name may go to a claimant that deserves it imperfectly. Loosely speaking, common sense is right. There are values, lots of them, and they are what we are disposed *de facto* to value.

Then is my position a form of realism about values?—Irrealism about values strictly speaking, realism about values loosely speaking. The former do not exist. The latter do.

What to make of the situation is mainly a matter of temperament. You can bang the drum about how philosophy has uncovered a terrible secret: there are no values! (Shock horror: no such thing as simultaneity! Nobody ever whistled while he worked!) You can shout it from the housetops—browbeating is oppression, the truth shall make you free.<sup>26</sup> Or you can think it better for public safety to keep quiet and hope people will go on as before. Or you can declare that there are no values, but that nevertheless it is legitimate—and not just expedient—for us to carry on with value-talk, since we can make it all go smoothly if we just give the name of value to claimants that don't quite deserve it. This would be a sort of quasi-realism, not the same as Blackburn's quasi-realism.<sup>27</sup> Or you can think it an empty question whether there are values: say what you please, speak strictly or loosely. When it comes to deserving a name, there's better and worse but who's to say how good is good enough? Or you can think it clear that the imperfect deservers of the name are good enough, but only just, and say that although there are values we are still terribly wrong about them. Or you can calmly say that value (like simultaneity) is not quite as some of us sometimes thought. Myself, I prefer the calm and conservative responses. But so far as the analysis of value goes, they're all much of a muchness.

<sup>26</sup> See Hinckfuss, *op. cit.*

<sup>27</sup> Simon Blackburn, *Spreading the Word: Groundings in the Philosophy of Language* (Oxford University Press, 1984), Chapter 6.



# DISPOSITIONAL THEORIES OF VALUE

Michael Smith, David Lewis and Mark Johnston

## III—Mark Johnston

### I

As a recently published letter reveals<sup>1</sup>, the same David Hume who seemed to suggest that distinctions of value were mere projections of our sentiments when he wrote<sup>2</sup>

Vice and virtue therefore may be compared to sounds, colours, heat and cold, which, according to the modern philosophy are not qualities in the object but perceptions in the mind.

also wrote

Philosophy scarce ever advances a greater paradox in the eyes of the people, than when it affirms that snow is neither cold nor white: fire hot nor red.

Taken together, and without prejudicing the interpretation of Hume, these quotations could serve as the motto of those who have attempted to defend a realism about value by way of an analogy with secondary qualities and with colour in particular. The aim of the analogists has been to undermine the characteristic claim of sentimentalist projectivism about value, namely that value is not a *genuine* feature of persons, acts, states of affairs, etc., but only appears so because we mistake features of our evaluative responses for features of such things. The leading idea of the analogists has been to show that by the same standards of genuineness it would follow that colour is not a genuine feature of surfaces.<sup>3</sup>

<sup>1</sup> Letter to Hugh Blair of 4 July, 1762, printed in *Mind*, October, 1986.

<sup>2</sup> *Treatise of Human Nature* edited by Selby-Bigge (Oxford, 1888), p. 469.

<sup>3</sup> See John McDowell 'Non-Cognitivism and Rule-Following' in Holtzman & Leich (eds.) *Wittgenstein: to Follow a Rule* (Routledge, 1981) and 'Values and Secondary Qualities' in Honderich (ed.) *Morality and Objectivity: a Tribute to J. L. Mackie* (Routledge, 1985). Also David Wiggins 'Truth, Invention and the Meaning of Life' and 'A Sensible Subjectivism' both in his *Needs, Values, Truth* (Basil Blackwell, 1987).

The analogy is opposed by Simon Blackburn 'Errors and the Phenomenology of Value' in Honderich (ed.) *op. cit.* and by Colin McGinn *The Subjective View* (Oxford

To fill out the leading idea: the conception of colour which the analogists typically employ is the so-called dispositional conception; according to which, for example

$x$  is red iff  $x$  is disposed to look such and so (ostended) way to standard perceivers as they actually are under standard conditions as they actually are.

Even if this biconditional misrepresents our conception of colour in general and of redness in particular, as I think it obviously does,<sup>4</sup> it is probably true that we could have employed a concept for which something like this was adequate. Operating with the fiction that such is our concept of redness, it first would follow that (categorically kosher) predications involving ' $x$  is red' have truth conditions and are straightforwardly evaluable as true or false. Secondly, redness would also be a genuine property: not simply in the sense that the predicate ' $x$  is red' and the abstract singular term 'redness' would have a semantic value, but also in the more interesting sense that predicating ' $x$  is red' of things could be part of straightforward casual/dispositional explanations of why those things look red to perceivers on *particular* occasions. (This latter should impress even a Nominalist who spurns the idiom of property-talk.) Thirdly, thanks to judicious placement of the rigidifying device 'actually', the biconditional allows that

---

University Press, 1983), Chapter 8. A sophisticated way of taking the analogy is presented by Crispin Wright in his 'Moral Values, Projection and Secondary Qualities' *Proceedings of the Aristotelian Society*, supp. vol. LXII, 1988.

Of course, the analogy is associated with a long tradition of 'moral sense' and 'ideal observer' theories. As far as I have been able to make out, it was Anthony Ashley Cooper, Third Earl of Shaftesbury, who in his *Characteristics of Men, Manner, Opinions, Times* (1711) first used the term 'moral sense' to denote aspects of our psychology plausibly taken to be attuned to value. He called the moral sense a 'reflex sense' which when applied to the objects of our affections gives rise to 'another kind of affection towards those very affections themselves', *Characteristics*, edited by John H. Robinson, 2 volumes, (Bobbs-Merrill, 1964) volume 1, p. 251. While it is unclear whether Shaftesbury intended, in Hume's crucial phrase, 'to derive moral distinctions from a moral sense', this psychologism was explicit in Francis Hutcheson's *An Inquiry into the Original of our Ideas of Beauty and Virtue* (London, R. Ware 1753), and in Adam Smith's *A Theory of the Moral Sentiments* (1759) where the ideal observer theory predominates.

A nice article on the tradition is James Ward Smith's 'The British Moralists and the Fallacy of Psychologism' *Journal of the History of Ideas*, XI, 1950.

<sup>4</sup>This claim is treated in some detail in my 'Objectivity Refigured' in *Realism and Reason*, John Haldane and Crispin Wright (eds.) (Oxford University Press, forthcoming).

something could have been red even if standard perceivers and standard conditions had been different and even if there were no standard perceivers and no standard conditions.<sup>5</sup> The dispositional view, so spelt out, is not simple subjectivism or idealism about colour. The colours of things are not existentially dependent upon our responses. Rather, colour concepts are conceptually dependent upon the concepts of our responses under certain conditions. (More on this conceptual dependence in section II).

Three kinds of theorists would be confounded if this were the correct account of 'red'. The tables would be turned on colour non-cognitivists, who deny that utterances of the form 'x is red' are truth-evaluable, instead supposing such utterances to be ejaculations merely prompted or caused by certain colour experiences under certain conditions. Given the dispositional account of 'red', some of the very conditions the non-cognitivist cites as the causal conditions for an ejaculation of the form 'x is red' turn out to be conditions under which the ejaculation is true.

Related difficulties arise for error theorists about colour, who suppose that while remarks of the form 'x is red' are or can be genuine assertions they are always in fact false, since nothing in the external world is coloured. The dispositional account shows immediately how external things could be red and indeed how we could get into a canonical condition for telling which things are red.

Finally, the account would expose as unnecessarily arduous the path of those delayed-reaction colour realists who, beginning with the non-cognitivist's starting point, see a problem about how remarks of the form 'x is red' could be truth-evaluable given their causal origin in our responses, a problem which requires a substantial explanation of how we come to 'earn the right' to express our experiences in terms of judgements about external things. On the dispositional account there is no *earning* the right, we have instead a natural conceptual right. The judgements about the colours of external things are judgements involving conditions on our experiences.

<sup>5</sup> On this use of 'actually' see Martin Davies and Lloyd Humberstone 'Two Notions of Necessity', *Philosophical Studies*, 1981.

Can the analogy between value and secondary qualities plausibly be deployed with corresponding effects against the non-cognitivism of A. J. Ayer, the error theory of J. L. Mackie and the quasi-realism of Simon Blackburn?<sup>6</sup> Everything depends upon just what analogy one has in mind.

Despite the fact that ordinary evaluators can immediately make evaluations on being perceptually confronted with complex situations, values are not in general the object of any perceptual or quasi-perceptual faculty or sense. As the one time currency of talk of the aesthetic sense or aesthetic attitude indicates, the sensuous aesthetic values are *prima facie* candidates for a perceptualist treatment. One can actually perceive the grace of a balletic movement, the satisfying resolution of a dissonance, the vividness of the depictions by the Sieneese School. To these we might add sensuous pleasures conceived as the objects of the various ordinary senses.

However, unless we are prepared to presuppose a substantive aestheticism or hedonism about value, we must admit the limitations of any kind of quasi-perceptual acquaintance in discerning the value of such things as truth, justice and the American Way. Cold old correspondence to the facts may leave us unmoved even if we are able somehow to vividly picture an isomorphism of truthbearer and truthmaker. This tells us not that truth is valueless, but that its value is not salient in such ways. Similarly with justice; for if justice involves deviating from equality in the distribution of power, resources or opportunity only when there is good and sufficient *reason* then anything that is like perception in presenting information via a depiction will be too crude an instrument by which to evaluate the justice of some distributional proposal. And of course, if there were such a thing as *the* American Way, it would be so multifarious in its details and so complex in its internal trade-offs that no depiction, perceptual or imaginative, could begin to discern its value. Any quasi-perceptual medium will thus tend to distort the evaluational message.

<sup>6</sup> See A. J. Ayer *Language, Truth and Logic* (Gollancz, London, 1936). A related non-cognitivist view is taken by C. L. Stevenson in *Ethics and Language* (New Haven, 1941). Mackie's error theory is set out in Chapter 1 of *Ethics: Inventing Right and Wrong* (Penguin, 1977), Simon Blackburn's quasi-realism is well presented in *Spreading the Word: Groundings in the Philosophy of Language* (Oxford University Press, 1984), Chapter 6.

A second and related point about the bruited analogy is that talk about red, colour or secondary qualities, while talk of determinables, is still relatively detailed talk. However, talk about value is talk at a level of almost fantastic abstraction. This remains so when we set to one side all so-called attributive uses of 'good' and talk of *a* good, and concentrate on the predicate 'x is a value' understood as a universal predicate of favourable assessment applicable to states of affairs. Any nominalization of any sentence has a state of affairs as its semantic value. So states of affairs can be relatively simple or complex, and valuable or not, in a variety of ways (aesthetically, hedonistically, morally, etc.) and for an enormous range of reasons. The only correlative condition of valuers with such enormous generality is the condition of their finding themselves with reason to value the state of affairs in question. That one is moved by a perceptual or quasi-perceptual representation of a state of affairs might indicate, in the absence of defeating considerations, that there is a reason to value it. However, it is incredible that being moved by such perceptions is the only sort of reason to value something or that all relevant reason-giving properly terminates in appeal to such perceptual promptings.

A third disanalogy between dispositional secondary quality concepts and the concept of value concerns the different prospects of an analysis or definitional reduction in the two cases. Colour terms, like terms for other secondary qualities, can be introduced by ostension. For example, we can rely upon the neophyte's quality space and a cannily chosen collection of foils and paradigms to make salient a way things look, going on to introduce the neophyte to a colour concept by saying: 'Something is red just in case it is such as to look the way these paradigms look from here now and not the way those foils look from here now, *but* look that way to standard perceivers under standard conditions'. If, in accord with our fiction that the orthodox dispositional account can be made to work, we could give a substantive specification of standard perceivers and standard conditions without even covertly using the notions of being red or being otherwise coloured, we would here have defined a colour concept ostensively and without relying upon any colour concept as opposed to colour sample.

However, nothing like this will be plausible in the case of the

concept of being a valuable state of affairs. Ostension cannot be relied upon here because there is no analogue of the perceptual quality space to secure stable generalizations from any set of foils and paradigms to the multifarious range of state of affairs which can be valuable. There is a colour (appearance) solid but no value (appearance) solid. This is a fundamental weakness in talk of a moral or evaluational sense. What is a sense without a quality space?

Fourthly, if, as suggested above, our finding reason to value something is a response of ours that is relevant to the thing's being valuable, then our finding *good* reason to value the thing is at least as relevant in any dispositional account of value. However the following looks like a trivial analytic connection: if and only if *x* is a *good reason* for finding state of affairs *y* valuable is it the case that if there is no countervailing reason, the state of affairs of valuing *y* because of *x* is *valuable*. Given this dependence of the notion of a good reason on the notion of value it appears that exploiting the notion of a good reason is giving up the analytically reductive game with respect to the universal predicate of favourable assessment of state of affairs. But isn't this to give up the only game suggested by the analogy between secondary quality concepts and the concept of value?

## II

No, it is not. The most plausible, if highly generalizing, way of taking the analogy is this: evaluational concepts, like secondary quality concepts as understood by the analogists, are 'response-dependent' concepts. Let me explain.

About many areas of discourse philosophers have urged a qualified realism, asserting both that the discourse in question serves up genuine candidates for truth and falsity, and that, nonetheless, the subject matter which makes statements of the discourse true or false is not wholly independent of the cognitive or affective responses of the speakers of the discourse. A basic problem of contemporary philosophy is the problem of explicating the relevant notion of dependence so that the qualified realism does not turn out to be (empirical) idealism in disguise, i.e., does not imply that the dependent subject matter would not exist or be the way it is but for the existence of our responses as they actually are. Elsewhere I have argued that neither Michael

Dummett's semantic anti-realism nor Hilary Putnam's internal realism are satisfactory solutions to the basic problem.<sup>7</sup> The approach I favour differs from these in not explicating the concept of truth in terms of the concept of our finding warrant to assert things under ideal epistemic conditions. Instead, a different conceptual dependence on concepts of our responses under specified conditions is exploited. The resultant qualified realism can be local and topic-specific, applying to some subject matters and not others, without it following that truth is an equivocal notion across different subject matters. Let us call those concepts which exhibit a conceptual dependence on or interdependence with concepts of our responses in certain specified conditions *response-dependent* concepts. How then are we to demarcate the response-dependent concepts?

If  $C$ , the concept associated with the predicate 'is  $C$ ', is a concept interdependent with or dependent upon concepts of certain subjects' responses under certain conditions then something of the following form will hold *a priori*

$x$  is  $C$  iff In  $K$ ,  $S$ s are disposed to produce  $x$ -directed response  $R$   
 (or  
 $x$  is such as to produce  $R$  in  $S$ s under conditions  $K$ .)

Moreover, for the concepts in question such a biconditional will not hold simply in virtue of a reading of  $K$ ,  $S$  or  $R$  which makes the biconditional trivial, imposing thereby no constraint on the concept  $C$ . Such a trivializing reading would be any reading which overtly or covertly specifies the conditions and subjects as whatever conditions and whatever subjects are required to get it right, or any reading which overtly or covertly specifies  $R$  as whatever response is truly  $C$ -detecting under the conditions specified. Given a 'whatever-it-takes' reading at any of these three points, the *a priori* truth of the biconditional so read indicates nothing in the way of the conceptual (inter)dependence in which we are interested.

However, when for a given  $C$  we have substantial or non-trivializing specifications of  $K$ ,  $R$  and the  $S$ s, and the resultant

<sup>7</sup>'Objectivity Refigured' *op. cit.*

biconditional holds *a priori*, then we have a concept interdependent with or dependent upon a concept of subjects' reactions under specified conditions. Such will be a response-dependent concept.<sup>8</sup>

Some concepts wear their response-dependent nature on their face: for example, the correlates of bodily sensation—the nauseating, the dizzying, etc.; the correlates of organ pleasure—the tasty, the titillating, etc.; the correlates of emotion—the shymaking, the embarrassing, etc.; the correlates of desire—the agreeable, the irritating, etc.; the correlates of belief—the plausible, the credible, etc.

The most obvious cases of *response-independent* concepts will be those concepts for which it is very plausible that the correct account of their content will not imply any substantial biconditional of the form above. Examples might be concepts of theoretical science such as the concept of a muon, logical concepts like conjunction, mathematical concepts such as successor. Of course a concept such as successor can be response-independent even if the concept of a subject's employing the concept of successor is response-dependent, being explicated in terms of the subject's dispositions to respond. Only a gross verificationism which confounded the conditions for possessing a concept with the content of the concept possessed would fail to recognize this possibility.

Many pivotal issues in philosophy, for example, issues about linguistic meaning, essence, personal identity, free will and the nature of similarity, can be cast in terms of whether and in what way the central concepts in those areas are response-dependent. So the central issue in the philosophical discussion of colour, the

<sup>8</sup> At least this holds with one proviso having to do with concepts introduced by reference-fixing descriptions and for which we have an *a priori* guarantee that there is some natural similarity underlying the relevant sample. Everyday terms for shapes might provide some examples. On reference-fixing see Saul Kripke *Naming and Necessity* (Harvard, 1980). For a discussion of the issue see 'Objectivity Refigured'. The distinction between response-dependent and response-independent concepts is one which I developed in my seminar on Ethics during the spring of 1986. I was fortunate enough to have Crispin Wright in attendance, and ever since then he and I have talked and corresponded about this and related distinctions. For his purposes he finds a different but related distinction useful. See his talk of 'order of determination' in 'Moral Values, Projection and Secondary Qualities' and his own contribution to the *Realism and Reason* volume. For a comparison of the distinctions see the appendix to my 'Objectivity Refigured'.



issues as between the so-called subjectivists and objectivists, is an issue over the response-dependence of colour concepts.<sup>9</sup> And the place to begin, though not to end, in explicating the primary/secondary quality distinction is with the response-independent, response-dependent distinction.

Notice that nothing in what I have said by way of characterizing response-dependent concepts implies that such concepts admit of a reductive definition or *analysis* in terms of concepts of subjects' responses. Hence the explicit allowance for conceptual interdependence.

The moral of recent philosophy is that many concepts, and most philosophically interesting concepts, have no interesting analysis. How can there fail to be an analysis of a concept *C* even although a substantial biconditional of the form above holds *a priori* for *C*? Well, it might be that in specifying the response in question we need to employ the concept *C*. The relevant *x*-directed response connected *a priori* to *x* being *C* might be the judgement or belief on the part of certain subjects under certain conditions that *x* is indeed *C*. Or it might be that in specifying the relevant conditions of response we need to require the stability of *x* with respect to precisely that range of determinables which includes the concept *C*. In the case of redness for example, we might want to rule out conditions in which the red thing would change colour if looked at. Or. . . .

That is, it may be that sometimes the biconditional of the relevant form which shows a concept to be response-dependent is strictly speaking circular. Circularity would be a vice if our aim were reductive definition. However our aim is not reductive definition but the exhibition of conceptual connections. In such an endeavour, circularity is a defect only if it implies the triviality of the biconditional. This is not the general case, for circular biconditionals of the relevant form are often sufficiently contentful to be open to *further* objection. (Remember, in the heyday of analysis, the critical papers pointing out that a proposed analysis was not only viciously circular but also subject to six counterexamples. The very nature of the charge indicates

<sup>9</sup>For some difficulties with attempts to characterize the debate between subjectivists and objectivists see the interesting papers by Peter Smith and Gregory McCulloch in the symposium 'Subjectivity and Colour Vision' *Proceedings of the Aristotelian Society supp. vol.*, LXI, 1987.

how a biconditional can be circular and contentful.) Indeed, when it comes to non-reductive explications, circularity can be a benefit and not all a defect, as we shall see. It would not be too strong to say that when it comes to explicating conceptual connections certain kinds of circularity are to be recommended.

Just to sketch the sense in which a response-dependent account of a concept *C* can be a qualified realism about instances of *C*, it should be obvious first that the holding of a substantial biconditional of the required form allows for a *realist* as opposed to *irrealist* conception of *Cs*, i.e., the account can allow that genuine instances of the concept *C* exist. So also, a response-dependent account of *C* is compatible with an (empirical) *realist* as opposed to (empirical) *idealist* conception of *Cs*; i.e. thanks to the dispositional formulation and the right sort of rigidifying on the actual responses under the actual conditions, the account can allow that the instances of *C* which do exist could still exist and be instances of *C* even if the relevant conditions and responders had not existed or had been different. So too, a response-dependent account of *C* can be a *realist* as opposed to an *anti-realist*, *internal realist* or *pragmatist* conception of *C*, i.e., it can deny both (a) that the meaning of statements employing the concept *C* is to be given in terms of those statements' conditions of warranted assertion, and (b) that the truth predicate applied to such statements is to be analyzed in epistemic terms. This latter is to say that one need not have a response-dependent account of the concept of truth in order to have a response-dependent account of some other concept *C*. (Indeed, there are reasons to think that any response-dependent account of truth will collapse into empirical idealism.)

In what sense then is a response-dependent realism essentially a *qualified* realism? Precisely in denying the independence of the concept in question from concepts of subjects' responses under specified conditions. Response-dependent realism is, if you like, a conceptual or transcendental idealism. However it implies no such things as noumena. (It may well imply that there must be available response-independent concepts of some things.)

The thesis I wish to explore is this: the notion of value, the all purpose notion of favourable assessment for states of affairs, is a response-dependent notion. The interest for present purposes of such a thesis is largely in the details. I have already indicated

that even if one thought that secondary quality concepts like the concept red were response-dependent one might reject the most obvious response-dependent account—the orthodox dispositional view. Similarly, I am not prepared to endorse just any response-dependent account of value. I begin with an elegantly straightforward account of value which, as it happens, is in my terms a response-dependent account. The difficulties for this elegantly straightforward account will lead naturally to my own preferred account.

### III

David Lewis claims that

- (L)  $x$  is a value iff we would be disposed to value  $x$  under conditions of the fullest imaginative acquaintance with  $x$ .<sup>10</sup>

We may take the things in the extension of the predicate here being analyzed to be propositions, including egocentric propositions, such as that *I* write a piece for the viola da gamba. We can think of such egocentric propositions in a variety of ways, e.g., as states of affairs under sentential characterizations which are egocentric. (Lewis has his own favoured way of dealing with so-called egocentric propositions. They turn out to be properties self-ascribed. However this difference *should* make no difference in what follows.)

Lewis is what I would call a response-dependent intuitionist about value! He writes

Compare a version of Intuitionism: by hard thought, one becomes imaginatively well acquainted with  $x$ ; in consequence, but not as the conclusion of any sort of inference, one intuits that  $x$  has a certain unanalyzable, non-natural property; and in consequence of that, one comes to value  $x$ . My story begins and ends the same. Only the middle is missing. Again an exercise of imaginative reason plays a crucial role. Again its relation to what follows is causal, and in no way inferential. But in my story, the consequent

<sup>10</sup> See his contribution to the present symposium, p. 113ff.

valuing is caused more directly, not via the detection of a particular property of *x*.<sup>11</sup>

Lewis tells us that (L) is not intended to indicate how we are to balance values but to determine what values there are to balance. Here, full imaginative acquaintance is the guide, according to Lewis. However, Lewis holds that there is not much chance of full imaginative acquaintance with all values at once. 'Imagination seems to involve simulation—getting into the skin of the part. How many skins can you get into all at once? Maybe tranquillity is a value, and so is vigorous activity; but maybe a full imaginative acquaintance with one precludes a full imaginative acquaintance with another.'<sup>12</sup>

After the values to be weighed have been determined, the balancer must, it seems, bring to bear some other technique besides or along with sheer imagination in order properly to weigh the values. There may well be a serious difficulty here. For many propositions, and so substituends for analysis L, describe evaluatively complex situations. Consider my living with Brünnhilde—a lively lass to be sure, but then there is that horse, those heroes and her family! There is no chance of adequately evaluating the proposition that I live with Brünnhilde without weighing values and disvalues. If imagination is the guide to value but not to the weighing of values then it can only be the guide to the value of evaluatively atomic propositions, i.e., propositions whose value is not compounded out of the value of more specific propositions. However this notion of a proposition being evaluatively atomic is rather obscure. It is not the notion of the *most* specific propositions. For the best candidates to be the most specific propositions, e.g., a measure of a physical parameter at a specific position, do not excite imaginative desire at all. Worse, whatever was plausible in talk of aesthetic value as organic unity suggests that it is utterly implausible to attempt to decompose the value of a painting or a symphony into a sum or any other function of the values of its constituent parts. That is, there is something heroic about the idea that we could build up to the value of a complex proposition, e.g., about a performance of a symphony, from

<sup>11</sup> *Ibid.*, p. 121.

<sup>12</sup> *Ibid.*, p. 126.

evaluatively atomic propositions, e.g. about particular notes.

On the other hand, if imagination were used as the means of balancing, we would not only have Lewis's problem of holding together in mind vivid awareness of conflicting values, but also the problem of biased weighting. When I consider the value of my refusing the bright baubles, fine wine and celestial dishes which have been offered to bribe me, the abstract appeal of justice might well get swamped. As Ogden Nash taught us, duty just does not have the visage of a sweetie or a cutie. Nor is it obvious that duty would acquire such a visage if only we were to more fully imagine it.

This last point may well be unfair to Lewis's intentions. Sometimes the direction 'Imagine more fully what  $x$  would be like' amounts to a direction to *think* more fully about  $x$ . Thinking more fully about the injustice of accepting a bribe and about why justice is the first virtue of institutional arrangements may well overcome the appeal of the baubles. And Lewis writes of 'imaginative reason'.<sup>13</sup> However, he imposes a strange constraint if thinking or reasoning to oneself is to be allowed. The relation of imaginative reason to the state of valuing is 'causal and in no way inferential'.<sup>14</sup> Indeed, the analogy with Intuitionism suggests that Lewis mainly thinks of imagination as a quasi-perceptual process. Reasoning and reflection may tell you what aspects of a complex situation to vividly imagine, but imagination, and not reflection in some broad sense, is to be the guide.

Whether or not it is Lewis's idea, the limitations of this idea are worth highlighting; for they point to the connection between value and substantive practical reason, a reflective capacity which subsumes but is not exhausted by *imagination in the narrow sense*—hereafter 'imaginative awareness'.

The fact that vivid imaginative awareness leads us to value something is at most a *prima facie* reason to consider it valuable. Critical reflection on the deliverances of imagination can and often should overturn patterns of evaluation which even very vivid and complete imaginative awareness prompts. Consider the habitually hopeful but hapless gambler who when he vividly remembers or imagines gambling is particularly moved by the

<sup>13</sup> *Ibid.*, p. 121.

<sup>14</sup> *Ibid.*, p. 121.

winning sequences he vividly recalls or imagines and not so moved by the losing sequences he equally vividly recalls or imagines. Although there is an ominous preponderance of losses, thanks to his optimistic outlook he is more encouraged by a win than discouraged by a loss. So his bookkeeping faculty fails him, and he still ends up valuing being a gambler. Critical reflection, here taking the form of some detailed tallying, is needed to reliably track the values. Nor is it plausible to suppose that as imaginative awareness gets *very* vivid and complete, correct tallying gets somehow built into it.

Moreover, in many cases, vivid and complete imaginative awareness may itself kill off legitimate valuation. Harmlessly frivolous activity, such as dressing up in unexpected costumes for a philosophy seminar, is a value and so legitimately valued. However it is of the nature of the value in the frivolous that it doesn't bear too much thinking upon, and certainly not very complete or vivid imagining. Represent it to yourself too completely or vividly and you may not be able to resist anticipating the embarrassment that would lead you to disvalue the strikingly frivolous. Yet so long as we are restricting ourselves to the harmless, when it comes to the frivolous the more striking, indeed the more surreal, the better. *Mutatis mutandis* for the erotic. Wouldn't one advise certain restrictions on awareness of the details of erotic goings-on? At least the practical wisdom of seduction is full of ways of clouding consciousness and partly masking reality.

Still another range of examples involves concern for others. Even if one is initially benevolent, complete awareness of the suffering of the mass of sentient beings would be horrifically depressing, and hardness of heart rather than valuing their release might well be the causal upshot.<sup>15</sup>

In each of these cases there is a sense of the value of the states of affairs in question which we have and which appears independent of our good guesses merely about the deliverances of vivid or complete imaginative awareness. What is the provenance of these convictions about value? I take it to be a kind of substantive practical reasoning about value. For many values we judge a

<sup>15</sup> For a similar point and some very telling examples and considerations see Alan Gibbard 'A Non-Cognitivist Analysis of Rationality in Action', *Social Theory and Practice* 9, 1983.

certain degree of awareness to be appropriate just because it is likely to encourage valuing what we independently take to be valuable on the strength of substantive practical reason.

That we rely on such relatively independent standards by which to judge the adequacy of the outputs of imaginative awareness is further shown by considering how Lewis's account stands with respect to the issue of the supervenience of the evaluative upon the descriptive. As Simon Blackburn has reminded us, the supervenience of the evaluative on the descriptive is an *a priori* or conceptual matter. Someone who claimed that he had discovered descriptive duplicates which differed in their evaluational features and not in virtue of extrinsic descriptive differences, such as a history of particularized attachment, would thereby show that he was not fully competent with evaluational terms.<sup>16</sup>

How is this *a priori* supervenience to be secured given L and Lewis's stipulation that the relation between imagination and the output of valuing is a purely causal one? The principle of *similar cause similar effect* is an empirical principle. We possess no *a priori* guarantee here. So we possess no *a priori* guarantee that if two states of affairs descriptively just alike were to produce duplicate feats of imagination in an appropriate subject then the causal output would be the same in each case. So given L and Lewis's stipulation, the *a priori* supervenience of the evaluative on the descriptive simply does not hold. G. E. Moore, who made so much of this supervenience thesis, avoids this problem as a direct result of his response-independent conception of value. His preferred process of intuiting value, like Lewis's, is causally related to valuing. However, for Moore, this causal process is in no way constitutive of the values of things, but is only the characteristic way of detecting those values.

The argument just given actually poses a problem for *any* purely psychologicistic account of value. It shows that the *a priori* supervenience of the evaluative will not be secured on any theory which allows that being a value has a sufficient condition stable purely in terms of a pattern of mere psychological causation, since it is conceptually possible that any such pattern can exhibit inconstancy in output for similar description-

<sup>16</sup> 'Moral Realism' in John Casey (ed.) *Morality and Moral Reasoning* (Methuen, 1971).

encoding inputs. Evaluative supervenience is instead grounded in a characteristic pattern of evaluative or practical reason-giving, a pattern to which the reason-giving must be answerable if it is to deserve the name of reason-giving. The practical reasoning which justifies adopting various evaluative attitudes towards states of affairs begins by taking into account and making something of the descriptive features of those states of affairs, just as inference to the best explanation begins by taking into account and making something of the observed phenomena. While with both sorts of reasoning different reasoners can make different things of the same material, part of what makes practical reasoning or inference to the best explanation deserve the name of reasoning or inference, as opposed to whim, fancy or imagination, is that the same constraints (as it happens not just deductive constraints) of *good* or correct reasoning or inference apply to all attempting to reason practically or inductively.

So suppose for illustrative simplicity that something like the following biconditional held *a priori*—

$x$  is a value iff practical reason is on the side of valuing  $x$ ,  
i.e., the deliverances of good practical reasoning support  
the conclusion that  $x$  is a value.

Now suppose also that  $x$  is a value, and let  $D_1, D_2 \dots$  be all of the descriptive features of  $x$  which practical reason would properly make something of in coming out in favour of  $x$ . Then any  $y$  alike to  $x$  in all descriptive features, and hence alike in  $D_1, D_2 \dots$ , would be similarly favoured by practical reason. So, if  $x$  and  $y$  are just alike descriptively then, if  $x$  is a value,  $y$  is also a value. The corresponding result is likewise derivable if practical reason counts  $x$  a disvalue. Nor need we think of practical reason as yielding for any  $x$  a single unequivocal result as to its value to derive supervenience in this way. Even if it is indeterminate whether  $x$  is a value and to what degree a value, in the sense that practical reason recognizes a number of tenable views on the matter, practical reason will yield the same range of outputs as acceptable or tenable for the same descriptive input.

Thus we derive and explain the *a priori* supervenience of the evaluative on the descriptive. It looks after all as if we can be evaluative realists and also treat the supervenience of the



evaluative on the descriptive as importantly unlike the empirically discovered supervenience of colour on primary qualities. What we haven't yet seen is how to be response-dependent evaluative realists and still do this.

The inconstancy of evaluative output from duplicate feats of imaginative awareness is not just a conceptual possibility showing that L lacks the resources to allow for supervenience. It is also an empirical possibility highlighting again the need for a constraint of reasonableness.

Imaginative familiarity with the same initially appealing material can breed contempt. But there is certainly room to ask whether the contempt familiarity breeds is warranted or unwarranted. Contrary to the implications of L, this is not to be decisively answered in terms of whether the later imaginative awareness was more vivid or complete than the earlier. It may be just that independently of variations in completeness or vividness the effect of the same imaginative material on subjects changes with repetition. Think of listening to the same simple and initially appealing piece of music many times over. After a while there need be no change in the degree to which I grasp the musical material. My interest simply wears off. There will then be no non-arbitrary way within the ambit of L to specify which outcomes of which repetitions are value-detecting and which represent the waning of my ability to imaginatively detect value in the material.

What counts as a defective response to value is not merely a matter of whether this or that psychological process occurs with this or that evaluative attitude resulting. What matters is whether the response in question is in accord with the deliverances of substantive practical reason about the subject matter in question.

The need for such a constraint of reasonableness on the actual and counterfactual responses which are to be counted crucial to the values of things points once again to the conceptual autonomy of normative notions, i.e., notions ordinary uses of which thereby express commendation or disapprobation. The concept of value cannot be analyzed or reductively defined by means of notions not themselves normative. Which is to say that there is no analysis of all normative notions. This conceptual autonomy of the normative is the counterpart for the normative

of Brentano's thesis about the conceptual irreducibility of intentional notions.

This pivotal thesis of conceptual autonomy is here intended to cover the so-called 'thin' ethical concepts, i.e., the concepts good, valuable, reasonable, worthy, *etc.*; concepts whose function is *essentially* to commend. Of course, the negative counterparts of these concepts, whose function is *essentially* to censure, are also included. Since I also hold that mortal danger *warrants* fear or that dishonesty *merits* censure it may seem that I should also extend the thesis of conceptual autonomy to the so-called 'thick' ethical concepts. However, I take it that we can say what both mortal danger and dishonesty are without using evaluative notions. More, the fact that dishonesty *merits* censure does not make it essential to the concepts of dishonesty that finding some person or act dishonest involves censure. All that follows is that under ordinary conditions there are substantive reasons on the side of responding to dishonesty with censure. We can imagine extraordinary conditions, e.g. of repressive and invasive political occupation, in which this is not so. We can also imagine ordinary circumstances in which further substantive reasons outweigh the reasons for censuring dishonesty—cases of 'white' lies if you like.

There are 'white' lies but no 'black' values. A full account of the essential connection between the content of the thin concepts and their commendatory function would explain facts like this. Here I offer to defend the claim of essential connection between the concept of value and its commendatory function by showing what this claim explains. The anti-reductive remarks above in connection with supervenience, plus the observations below about the 'So What?' argument, about the iterativity of value and about relativism point to various manifestations of the essential connection.

#### IV

Speaking now quite generally of accounts of the concept of value, suppose that a theorist proposes that the valuable is what we would value under condition *K*, where *K* is not an evaluatively or normatively characterized condition. Someone informed of this putative analysis might now employ the *analysans* instead of the *analysandum*. But would they be

employable to the same effect? To say that something is valuable is thereby essentially to commend it. To say that something would be valued under condition *K* is not thereby to commend it, but only to make a descriptive remark about its relation to certain psychological conditions. The commendatory function of the original remark would be recaptured if one said not only that it would be valued under condition *K* but also that *K* is the right or a reasonable condition to be in when considering questions of value. This is of course to reimport normative notions. Furthermore, since 'is valuable' is supposed to be the most general term of commendation, and since being right or being reasonable in the context under discussion are ways of being valuable, it looks as though there will be no reductive definition of 'is valuable' available at all. So there will be no reductive response-dependent account of 'is valuable' that is tenable.

Another manifestation of the conceptual autonomy of the normative is the so-called 'So What?' argument. Gilbert Harman, for example, applies the argument against the ideal-observer version of a response-dependent account of value.<sup>17</sup> That version has it that to be a value is to be such as to be valued by a certain kind of observer. Calling the observer 'ideal', although part of the tradition, can be misleading. For 'ideal' is here supposed to be a stand-in for a purely descriptive characterization of the kind of observer in question. Now suppose that I am told that while I value pushpin the 'ideal' observer thinks it worth nothing. I infer that the conditions under which I make my evaluation of pushpin are not the conditions in which the 'ideal' observer makes his evaluations. Notice that I may think this without in any way thinking worse of my evaluation. Upon being told the difference between my evaluation and that of the 'ideal' observer, I may say, without showing any contempt for reason or value, 'So What?'. That is, I may coherently take it that there is no reason to get into the 'ideal' condition and no reason to correct my valuations to accord with my beliefs about the evaluations I or another would make in those conditions.

<sup>17</sup> 'Moral Agent and Moral Spectator', *The Lindley Lecture* (Kansas University Press, 1983).

An ardent defender of the ‘ideal’ observer analysis might now insist that there is a reason; namely that, as his analysis shows, the ‘ideal’ conditions are those in which the real values present themselves. Although this is the consistent thing for the ardent defender to say, it is not a very plausible thing to say just because the conditions cited are not normatively or evaluatively loaded. That is, it will not be very plausible just because saying that something would be valued by the ‘ideal’ observer is not thereby to commend it. If this is the right account of the force of the ‘So What?’ argument then the argument holds against any reductive account of value.

A third and more novel way of bringing out the conceptual autonomy of the normative begins with the observation that the following principle has an *a priori* status. I mean we know it, but not on the basis of any empirical knowledge.

#### Iterativity of Value

Among the things that are valuable this is:  
to value the valuable.

Now part of the point either of giving an analysis of the concept of value or of giving central play to an *a priori* biconditional of the form

$x$  is a value iff  $S$  would value  $x$  under condition  $K$

is to articulate a conceptual connection which is relatively basic or generative in the sense that some of the central *a priori* truths about the relevant subject matter are derivable from it, perhaps in conjunction with other relatively basic *a priori* truths. For the derivation of the iterativity of value to work, something like the following is needed:

Psychological state  $K$  is such that in  $K$  one would value valuing what one would value in  $K$ .

This claim and the biconditional will yield the iterativity of value. For all I know the claim may be true on certain specific ways of spelling out the psychological state  $K$  and the nature of the psychological state of valuing. However any such truth would be true as a matter of empirical psychology. Any derivation employing such a truth would not secure the iterativity of value as an *a priori* matter.

How could such a result be secured? One response is to claim that it cannot be secured and hence that it is nothing against a proposal that it does not generate iterativity. However notice that the following principle also seems *a priori*—the iterativity of what reason supports.

Among the things reason is on the side of valuing is this:  
valuing what reason is on the side of valuing.

If we could connect up value with what reason is on the side of valuing, we could derive the iterativity of value. Of course something will remain underived, and one would have to use one's philosophical judgement as to whether the iterativity of value was interestingly derived from the iterativity of reason and the connection between reason and value. Myself I think it is. But for present purposes all we need is that if the iterativity of value is interestingly so derived then one has a further motive for exploring the conceptual connections between the normative notions of value and reason. As to the connection between something being a reason and our taking it to be so, more later.

So far then, we have argued against L on account of the limits of imagination, and against a whole class of analyses like L on account of their conceptually reductive character.

Lewis proposes a conceptual reduction at another point. He suggests that the crucial attitude of valuing is just desiring to desire. Here 'desire' is not to be understood as a state which is coherent only in the believed absence or non-possession of its object. Someone who knows he has a million can desire that million in this sense. The term 'desire' is functioning for Lewis as an all purpose term for any first-order pro-attitude.

I doubt the identification of valuing with desiring to desire, not only because of cases in which one desires to desire without valuing<sup>18</sup>, but also because of cases in which one values and yet has reason not to desire, and hence not to desire to desire, reasons which might be thoroughly effective. One case of the latter sort has to do with satisfactions which we might call

<sup>18</sup> Discussion of such cases would require examination of Lewis's distinction between instrumental and intrinsic desire and of the extent to which this very distinction is conceptually dependent upon the distinction between instrumental and intrinsic value. See Lewis, footnote 4.

*strongly serendipitous*, i.e., satisfactions which as a matter of psychological fact or artificial circumstance come only to those who are without any prior desire for them. To those who want such satisfaction they shall not be given. Knowing that some satisfaction was strongly serendipitous in this sense, I could reasonably and without doublemindedness value it and precisely not want to want it. Notice that this kind of point could be made at any place in the hierarchy of higher-order desires, throwing into question the identification of valuing with any order of iterated desire. Such considerations suggest that the attitude of valuing may be none other than the attitude of judging valuable. Here we have a further insult to reductionist sensibilities.

Part of Lewis's motive for the identification of valuing with desiring to desire is to secure something in the way of Internalism, i.e., a conceptual connection between value and motivation. Lewis explains the connection in this way—'If something is a value, and if someone is one of the appropriate 'we', and if he is in ideal conditions, then it follows that he will value it. And if he values it, and if he desires as he desires to desire, then he will desire it.'<sup>19</sup> The connection is not only very iffy, as Lewis himself points out. It is also a connection whose holding secures no privileged relation between value and motivation. For consider the *alcovalues*; where *x* is an alcovalue just in case we would desire to desire it when plootered, i.e., under conditions of extreme drunkenness. If something is an alcovalue, and if someone is one of the appropriate 'we', and if he is in ideal alcoholic conditions, then it follows that he will desire to desire it. And if he desires to desires it, and if he desires as he desires to desire, then he will desire it. Surely however, Internalists meant to articulate a *privileged* relation between *value* and motivation; in fact, a relation between judging something valuable and being motivated to pursue, cherish or maintain it, a relation supposed to hold whether one made the judgement of value in canonical conditions for valuing or accepted it on the say-so of one's spiritual advisor.

The prospects for deriving from any response-dependent account of value *this* non-contingent and direct relation between

<sup>19</sup> Lewis, p. 117.

judging something valuable and being positively motivated towards it seem to me dim.<sup>20</sup> But this is no objection to a response-dependent account of value. There is a name for not being motivated by what one judges valuable. It is of course 'weakness of will' and, so far as I can see, weakness of will not only can disrupt the expected connection between judging something best and acting to promote it but also can disrupt the expected connection between judging something valuable and desiring it in the extended sense.

As for securing an internal or conceptual connection between value and the will, *this* at least is true: to the extent that one is not weak-willed one will desire (in the extended sense) as one judges valuable. So much is part of the definition of weakness of will. As far as making the connection between judging valuable and desiring (in the extended sense) particularly intelligible, this seems to me achieved by the observation that 'valuable' and 'desire-worthy' are near synonyms. If judging valuable is pretty much judging desire-worthy then it is readily intelligible why judging valuable should lead to desiring. As the 'So What?' argument indicated, the real difficulty is to preserve this ready intelligibility *after* giving a philosophical gloss on the concept of value or desire-worthiness.

## V

David Hume also wrote 'Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger'<sup>21</sup>; to which an adequate response might be: not contrary to reason in one sense perhaps, but brutally insane, psychopathically callous and demonically indifferent. The widespread appropriation of the terms 'reason' and 'rational' for (i) logical, mathematical and perhaps probabilistic consistency among beliefs, combined with (ii) a tendency to maximise utility in the decision theoretic sense, may be ideologically suspect, but it is probably not bad English. Allowing the appropriation, we should also be mindful of that more substantive reasonableness

<sup>20</sup> But see Michael Smith's contribution to this symposium.

<sup>21</sup> *Treatise*, 'On the Influencing Motives of the Will'.

which we look for in both practical and theoretical matters. One who infers to the best explanation of phenomena, which involves more than maintaining probabilistic consistency in the face of the phenomena, is substantively reasonable, as is one who jumps into a swimming pool to save his drowning father-in-law even though it will mean getting his pants wet.

Of course, while those who share our form of life share a rich body of belief about what is substantively reasonable, important aspects of what is substantively reasonable are essentially contested. That is just to say that many questions of value are essentially contested. For what is typically not in question among the contestants is that

- (1)  $x$  is a value iff substantive reason is on the side of valuing  $x$ .

The concepts of value and of substantive practical reason take in each other's washing. The point of this Austinian metaphor is to suggest that (1) may provide no interesting analytic account of value but only a paraphrase. It is as well however to start by enshrining acceptable paraphrases, not only because many putative analyses simply fail to comport with or capture them, but also for the now familiar reason that analysis may not be a reasonable aim.

We arrive at something a little more controversial if we add to (1)—

- (2)  $y$  is a substantive reason for/against valuing  $x$  iff we are disposed stably to take it to be so under conditions of increasing information and critical reflection.

Increasing information will come at some point to include the revelations of imaginative awareness about the relation between  $x$  and  $y$ . But it will also include information from other sources, as well as information about the effect of this relation on imagination, and the effects of such imagination in its turn. Critical reflection will then involve attempting to debunk the beliefs which result from the informational input by examining the extent to which those beliefs are the result of prejudice, illusions of salience, self-deception, sour-grapes, false consciousness or other self-protective attitudes which defeat the prima



facie claim of informed belief to be true belief. The undebunked beliefs, still beliefs bearing upon the reasonableness or unreasonableness of valuing  $x$  because of  $y$ , are then to be brought into a wider reflective equilibrium with the rest of what we believe.<sup>22</sup>

The requirement that the output of such a process be stable under increasing information will be satisfied just in case (i) there is some state of relevant information about the issue from which critical reflection yields the output, and (ii) any application of critical reflection to *more* inclusive states of information would yield the same output. This slightly round-about formulation is needed to secure the result that although there may be no ideally informed critical condition, the real reasons are those that stand, and would continue to stand, the test of criticism.<sup>23</sup>

Along with (2), we need a clause telling us what the acceptable methods of weighing the various reasons for and against valuing  $x$  might be. Clearly, we should allow our beliefs about methods of weighing reasons to be subject to the same potentially mutating influence of critical reflection. Thus, mirroring (2), we have—

- (3) A method for weighing substantive reasons is an acceptable method for determining whether the weight of substantive reason is on the side of valuing  $x$  iff we are disposed stably to take it to be so under conditions of increasing information and critical reflection.

Now we say

- (4) Substantive reason is on the side of valuing  $x$  iff this is so

<sup>22</sup> On debunking, see Susan Hurley *Natural Reasons* (Oxford University Press, 1989). A crucial type of attempted debunking, relevant to the holding of supervenience, is testing our reactions by the requirement that like cases be treated alike. Thus, critical reflection as characterized here secures the supervenience of the reasonableness of acts of valuing on the descriptive features of those acts. That is, if  $w$  and  $z$  are like cases, acts of valuing descriptively just alike, then, if  $y$  is a reason for  $w$  it is also a reason for  $z$ . (I am, by the way, indebted to Susan Hurley for extensive discussion of this and many other issues surrounding this paper.)

<sup>23</sup> A fuller discussion of these matters would have to take into account the relationship between J. David Velleman's penetrating critique of Richard Brandt's theories of ideal rationality in his 'Brandt's Definition of the Good', *Philosophical Review* XCVII, 1988 and the present theory of rationality.

according to one and all methods of weighing the reasons for and against valuing  $x$ .

Taking (1) through (4) together, the response-dependent element in the notion of value will be evident. The notion of value is conceptually connected to the notion of substantive reason, as (1) makes clear. The notion of substantive reason is response-dependent as (2) and (3) indicate. That is, (2) and (3) together imply that substantive reasonableness is not hyper-objective or, reworking terminology of Bernard Williams, *hyper-external*.<sup>24</sup> There are no substantive reasons which we cannot get to in principle from here, although getting to them may involve a gradual but thorough reworking of what we take to be substantive reasons, the appropriate methods of weighing them and perhaps also the correct styles of critical reflection. If we think of our present system of substantive reasonableness on the model of Neurath's ship, not only may the ship require considerable overhaul but so also may our methods of overhauling it.

Clause (4) requires some comment. Allowing that there may be different acceptable ways of weighing reasons seems necessary given the enormous diversity of value, the controversial nature of claims about how to weigh evaluative considerations from disparate domains, and reasonable skepticism about critical reflection's capacity to dissolve such controversy. So there are moral considerations, considerations deriving from personal ideals and loyalties, considerations of utility, aesthetic considerations, and so on. We can only speak non-misleadingly of  $x$  being a value *simpliciter* if the considerations within and among these disparate domains tend in the same direction when subsumed under *any* acceptable method of weighing. Otherwise we should issue a divided report on the value of  $x$ , qualifying our talk of its value in terms of the various kinds of disparate considerations and methods of weighing which support or count against valuing it.

Divided reports will also be necessary whenever the considerations within or across evaluative domains are strongly

<sup>24</sup> 'Internal and External Reasons' in B. A. O. Williams *Moral Luck* (Cambridge, 1984). Williams doubts the existence of external reasons even in the sense defended here.

incommensurable, allowing for no acceptable method of weighing.<sup>25</sup> (Hence the rider 'one and all' in (4). We want to cancel the usual logical implication of the vacuous case.)

Strong incommensurability aside, one only has to think that substantive reason enforces no single method of weighing considerations to regard as conceptually amiss the consequentialist idea of maximizing value *simpliciter*, producing the most valuable or maximally valuable total state of affairs. Divided reports will be too common for this to be a reasonable aim. This suggests that there is more connection between G. E. Moore's response-independent conception of value and his Utilitarianism than is usually recognised. Once one takes a response-dependent view of value, one cannot reasonably hope that the disparate and divided nature of valuation, relatively robust as it is even under conditions of critical reflection, simply testifies to our ignorance of the true values and their true orderings.

Of course, the need for divided reports and the resultant conceptual difficulties with consequentialism would be much lessened if there were a ubiquitous kind of consideration which itself allowed for a privileged commensuration and which trumped other reasons.

The idea that considerations of utility (desire-satisfaction) are of this sort seems to me just a horrible and obvious evaluative error. So I am not the one to give this idea a run for its money. On the other hand, there is the idea, made somewhat plausible by the seriousness of moral considerations, that moral considerations are overriding. At various times and places the social institution of morality has encouraged this pretension by connecting moral reasons with a metaphysical picture of value as unitary and response-independent—as it might be, our true and only end is union with God, and an act is right only to the extent that it makes this more likely. Even Kant's wonderful claim that the only thing in this world or in any other that is good-in-itself is a good, hence autonomous, hence moral, will seem to me to rely upon a response-independent conception of value. For as soon as being good-in-itself is understood in terms

<sup>25</sup> On the complex issues surrounding incommensurability and on strong forms of it see Michael Stocker *Plural and Conflicting Values* (Oxford University Press, 1989).

of our finding non-derivative reason to value the thing in question, it becomes clear that more than the moral will qualifies and that the goodness of the moral will need not be overriding on all methods of weighing reasons which we find acceptable.

Hence, on the response-dependent view of value it may be slightly mistaken to ask: how much does morality require of me in the way of sacrificing my personal ideals? Within limits, substantive reason may recognize a range of equally acceptable ways of striking the angels' bargain. Within those limits, the issue may be a matter of invention rather than discovery.

## VI

There is an important distinction between allowing the sensible idea of reasonable alternatives within any idealization of our scheme of substantive reason and allowing the probably incoherent idea of reasonable alternatives without. If you like, this is the distinction between Pluralism and Relativism. Relativism can mean the discovery of a surprising relativity in our concepts. In that sense any response-dependent account of value is relativistic. But here I fall in with the tradition of use which appropriates 'Relativism' for a conceptually risky mix of adherence to standards *and* bland tolerance towards those who flout them, so long as the flouters are appropriately alien.

The Relativist begins by pointing to the possibility of the Radically Other—beings who share nothing of our sense of substantive reasonableness though they be perfectly rational in the decision-theoretic sense. Rational crabs perhaps? Despite the exercises of the subtlest philosophy, I see no incoherence in the Relativist's claim that such a radical alternative is possible.<sup>26</sup>

Incoherence threatens however when the Relativist tries to persuade us that there is really nothing to choose between our system of evaluation and the aliens' system of quasi-evaluations. The Relativist's route to this result invariably involves relying upon a neutral viewpoint allegedly incorporating all the information about both patterns of response, their causal histories and analogous roles in the lives of the two communities.

<sup>26</sup> I have in mind Donald Davidson's 'On the Very Idea of a Conceptual Scheme' in his *Inquiries into Truth and Interpretation*. In fact it seems to me that this paper contains the premises for a proof that alternative conceptual schemes *must* be possible.

From this viewpoint, it is claimed, no privilege can be seen to attach to our way of going. Notice that this is not to say that the Relativist requires us to make at the neutral viewpoint a judgement of equivalence of value. That would be straightforwardly inconsistent, inviting us to value in terms not our own. Instead the Relativist must restrict himself to trying to undermine any preference we would have at the neutral viewpoint, showing it to be arbitrary in the light of the analysis of value and full information about the two ways of going.

Everything depends on what is involved in the neutral viewpoint. One escape route from incoherence is to find in the neutral viewpoint a higher-order evaluative system, realized both in one's own pattern of response and in the aliens' pattern. However this is also an escape route from Relativism to a higher-order Absolutism. To remain Relativistic and make sense of the neutral but *fully informed* viewpoint, the Relativist must, I think, rely upon descriptive reductionism about evaluative and quasi-evaluative judgements, talking as if judgements of each sort could be analyzed in descriptive terms *without evaluative remainder*.

For example, it might be that although about as vulnerable as we are, the aliens are naturally self-destructive in the face of danger. So while it is true that

$x$  is mortally dangerous iff  $x$  is such as to produce the response of mortal fear in us under ordinary conditions.

it is also true that

$x$  is mortally dangerous iff  $x$  is such as to produce the 'in for a penny, in for a pound' response in them under ordinary conditions.

Here we have two opposing and, in one sense, equally natural dispositional patterns of response in the face of mortal danger. For every fact about our underwriting our pattern of response, there is a fact about their underwriting their pattern of response. There is nothing in nature to favour the one pattern of response rather than the other unless we suppose nature somehow gives special value to survival. But the Relativist tells us that all the relevant facts here merely involve our being disposed to desire to desire survival more than they desire to desire it. Taking into account all this information and remaining rational in the

formal sense, we will find ourselves without the resources to judge their ways to be worse than ours.

*Notice that the Relativist is here trying to invest the illusory capital of the 'So What?' argument.* Whereas we earlier employed this argument to show that no reductive account of value is correct, the Relativist employs a reductive account of particular values (or as in the case of mortal danger, disvalues) in order to motivate a viewpoint from which the difference between two patterns of response leaves us with the 'So What?' reaction generalized.

However, once we have abandoned definitional descriptive reductionism about value we can see that the Relativist has indeed omitted some crucial information from the neutral viewpoint, namely that substantive reason is on the side of (improvements of) one pattern of response and not the other. This is not to say that this information is in principle enough to persuade even the aliens that our ways are better. It is to say that, given this information, *we* have decisive and, for us, typically motivating reasons to favour the one pattern of response rather than the other. For, as we saw under the heading of iterativity, substantive reason is on the side of valuing what substantive reason sides with.

So for example, what the Relativist misses in the little story about mortal danger is that it is not just that mortal danger happens to *produce* fear in beings like us. Mortal danger *warrants* fear, i.e., in the ordinary run of things, substantive reason is on the side of feeling fear in the face of mortal danger. What is vulgar in vulgar Relativism is precisely the assumption that such commendations of our patterns of response will get captured or replaced by purely descriptive analyses. A non-reductionist about value can thus hold out against Relativism, even granting the possibility of the merely quasi-evaluating aliens.

For the Relativist's allegedly neutral viewpoint is one from which we are supposed to see any tendency to favour our own way of going on as a failure to analyze that tendency as just another disposition fully characterizable in descriptive terms. As we said, this constant undermining, rather than an attempt to get us to make a global judgement of equal *worth* as between the two ways of going, has to be the Relativist's strategy, otherwise he is simply inconsistent, inviting us to value in terms not our own. But the undermining strategy also fails. For our tendency

to favour our own pattern of response, or rather idealizations of it, is warranted, i.e., supported by substantive reasonableness, and this in its turn is a matter of our finding certain considerations substantively reasonable under conditions of increasing information and critical reflection. These facts are not neutral in the required sense. For our finding certain considerations substantively reasonable warrants our taking those considerations and not others into account. Given such a non-reductive story about our evaluations, the Relativist's viewpoint cannot be both neutral *and* fully informed. Formal rationality augmented with a host of non-commendatory descriptions of our responses falls short of full information about those responses. (Here we individuate information in terms of *a priori* equivalences, as seems independently plausible.)

This discussion of Relativism allows for the possibility of a 'they' disjoint from the 'we' of the present account, because the subtlest philosophy will probably continue to fail to show that there could be no such 'they'. On the other hand, the most informed anthropology is complexly related to the empirical question of the extension of 'we'. For we have no evidence of a 'they' disjoint from us whenever the diversity of valuation is to be explained by false consciousness, ignorance of fact, failure of criticism, or the adoption of divergent mores in reasonable response to divergent histories and contemporary conditions.

Starting with myself and my soulmates, the extension of 'we' in the present response-dependent account is to be determined by including with us all who would stably converge on the same judgements of substantive reasonableness under increasing information and critical reflection.<sup>27</sup> 'We' is in this way ideally inclusive. For all I know it may include some of the more savoury characters of the Marquis de Sade's *Juliette* and a few of Bernard Williams' samurai to boot.<sup>28</sup>

Despite this ideal inclusiveness of 'we', the resultant account is not trivial or empty. For what I (and my soulmates) would take

<sup>27</sup> Exactly how much convergence is required? We can leave this vague, in effect offering a cluster of accounts instead of a single one. We want each account within the cluster to provide a reasonable trade-off between the extent of our solidarity and the extent of our agreement. See Lewis's remarks around his footnote 20, and his 'Scorekeeping in a Language Game', *Journal of Philosophical Logic* 8, 1979.

<sup>28</sup> For the samurai see 'The Truth in Relativism' in B. A. O. Williams, *op. cit.*

to be a reason under the ideal conditions is a matter fixed (within a range of indeterminacy) by my psychology and the details of the idealization.

The account may now seem shamelessly self-congratulatory. For it implies that there is a conceptual guarantee that I and my soulmates, or to put the point more bluntly, I, could not be invincibly immune to reason. (After all, just who counts as a soulmate gets determined by empirical facts and the inclusive criterion.) That is, thanks to the response-dependent account of value and the present gloss on 'we', I can know *a priori* that I myself would not fail to track substantive reason under the ideal conditions. However, sad to say, the corresponding truth about you, if it is a truth, is an empirical truth.

This is an objection. But I think that the best that a response-dependent theorist of value can do is the best that can be done. Any residual oddity is testimony to the intuitive residue of a kind of projective or hyper-objectifying error about reason and value.

The best I can do in response to the objection is to point out that you can similarly congratulate yourself. Although there is a strict sense of predicating in which we do not predicate the very same thing of ourselves when we each congratulate ourselves as not invincibly immune to reason or value, there is nonetheless every chance that we will be talking about pretty much the same feature in extension. Value from my point of view may massively overlap with value from your point of view.

## VII

The idea that talk of value and reason is unequivocal, quite determinate, and not equivalent to any talk about what we would count valuable or reasonable is entrenched and naturally makes for some residual dissatisfaction with any response-dependent account. But a response-dependent account of value or reason should be offered as a partly revisionary account.

In the *Lecture on Ethics* Ludwig Wittgenstein wrote

The right road is the road which leads to an arbitrarily predetermined end and it is quite clear to us all that there is no sense in talking about the right road apart from such a predetermined goal.



. . . I think it would have to be the road which *everybody* on seeing it would, *with logical necessity* have to go or feel ashamed for not going.<sup>29</sup>

There is something in the idea that if anything is absolutely and strictly to deserve the name of value it would be a practical demand built into the world in such a way that any merely formally rational being would on pain of inconsistency feel shame if he were not to respect it. This is in effect what John Mackie called the idea of value as 'the objectively prescriptive', an idea which Mackie successfully stigmatized as an error at the heart of our thought about value. Mackie offered a projective metaphor as an account of the provenance of this hyper-objectifying error. According to Mackie, the error arises because we mistake the felt urgency of reasonable response for the pull of a demand thoroughly independent of that tendency to respond. Whatever the origin of the error, the present response-dependent account of value is designed to eliminate precisely this error of supposing that the demands of value or substantive practical reason are thoroughly independent of our tendency to respond to such demands.

If the error is as deeply rooted in our thought about value as Mackie suggests then we should expect recurrent manifestations of the error. Something related to the error is elegantly enshrined in Plato's *Euthyphro*. There Socrates objects to Euthyphro's definition of the pious as what the gods love. Socrates insists that rather than acts being pious because they are loved by the gods, the gods love such acts because they are pious.<sup>30</sup> It simply turns out that, thanks to their natures, the gods are very good detectors of piety which in fact they invariably love. There is, as we might put it, nothing *a priori* about this, as is *shown* by the explanatoriness of the remark 'The gods love pious acts because they are pious' as compared with the lack of explanatoriness of the remark 'The gods love pious acts because they are the acts the gods love'. If *a priori* equivalents can be substituted in explanatory contexts saving explanatoriness, then Euthyphro's account fails, even taken as a mere allegation of *a priori* equivalence.

<sup>29</sup> Published in *The Philosophical Review* LXXIV, 1965.

<sup>30</sup> *Euthyphro*, 10B-11B.

This is precisely the form of argument which highlights the felt independence of value from value-directed responses. It is the characteristic bugbear of response-dependent accounts. Notice however that our kind of response-dependent theory of value will allow many instances of this form of argument as showing that *certain* dependencies are not relevant to the correct account of value. Socrates does indeed have the better of Euthyphro. The fact that the gods love certain acts is not central to the account of piety. Similarly with the ideal observer account of value. Something is valued by the ideal observer because it is valuable, i.e., in accord with substantive practical reason. It is not valuable because it is valued by the ideal observer.

However when it comes to this instance—something is not a reason because we take it to be so, we take it to be so because it is a reason—care must be exercised. Our kind of response-dependent theory can discern an explanatory element in the remark that we take something to be a reason because it is a reason. To use an analogy suggested by Michael Slote, just as we can explain the behaviour of a gas in terms of the way it would behave if it were ideal, we can explain our reason-accepting behaviour in non-ideal conditions in terms of our there approximating what we would accept in ideal conditions.<sup>31</sup>

But now the Socratic objector will take the decisive step, insisting that whatever is a reason is not a reason because we would stably take it to be so as we approach ideal conditions. Rather, in the most fortunate case, we would take it to be a reason under such conditions just because it is a reason. And now our kind of response-dependent theorist must dig in. The hyper-objectifying error has at last been manifested. The objector in effect wants hyper-external reasons, reasons which could in principle outrun any tendency of ours to accept them as reasons, even under conditions of increasing information and critical reflection.

We should now ask the Socratic objector by what consideration he is led to believe that we have ever been in contact with the true hyper-external reasons, or better, by what consideration he

<sup>31</sup> Michael Slote 'The Rationality of Aesthetic Value Judgements', *Journal of Philosophy* 68, 1971.

is led to believe that increasing information and critical reflection is a good way of seeking such contact.

For the Socratic objector must regard

- (H) The substantive reasons are to be discovered by taking into account relevant information and critically reflecting on it.

as a contingent empirical hypothesis. But how could such a hypothesis be confirmed? After all we have no independent grip of the idea of the substantively reasonable besides the idea of what is discovered by informed criticism. So there is something essentially bogus about the very idea of finding empirical support for the connection of substantive reasons with such discoveries. At this point, I think the descendants of Euthyphro get the better of the descendants of Socrates.

I cannot expand on these considerations here, but as they stand they may help to locate the present response-dependent account with respect to Simon Blackburn's quasi-realism.<sup>32</sup>

Blackburn himself has provided a battery of persuasive arguments against any analysis of the truth conditions of evaluative judgements in non-normative terms, including versions of the 'So What?' argument and the Euthyphro argument. He has also quite rightly challenged evaluational realists to explain how they propose to generate the *a priori* supervenience of the evaluative upon the descriptive. His own view is that no interesting analysis is available and that the philosophical task in the theory of value as elsewhere is to explain *how*, given the essentially non-cognitive attitudes (sentiments, concerns, etc.) which are expressed in evaluative discourse, we could earn the right to express those attitudes in truth-conditional form.

The non-reductive response-dependent account of value can take on board much of the polemic against reductive accounts of value. But it also implies that there is no need to *earn* the right to the truth-conditional form of expression. Modulo the correction of a hyper-objectifying error, we have a natural conceptual right to this truth-conditional form. Our response-dependent account

<sup>32</sup> See Chapter 6 of *Spreading the Word*; 'Morals and Modals' in *Truth, Fact and Value*, Graham Macdonald and Crispin Wright (eds.) (Oxford, 1986); 'How to be an Ethical Anti-Realist' in *Midwest Studies in Philosophy*. Volume XII, Peter French et al. (eds.) (Minnesota, 1988).

of the concept of value does not analyze truth-conditions, it rather explicates them sufficiently to exhibit the interconnections with our sensibilities and concerns. Moreover, as we saw, the response-dependent account has the resources to explain the *a priori* supervenience of the evaluative on the descriptive.

An appropriate response-dependent account may thus threaten to make quasi-realism redundant. The quasi-realist programme is to *somehow* defend our right to employ the truth-conditional idiom in expressing evaluations. But what kind of truth-conditional idiom? An idiom whose central concepts are response-dependent or an idiom whose central concepts are response-independent? If the former, then quasi-realism might as well straight-out endorse its preferred response-dependent account. If the latter, then quasi-realism is not redundant, but it faces the difficult task of explaining how, given the concern-expressing point of our evaluational talk, we end up rightfully employing truth-conditions whose contents appear to transcend that point. If anything like the present account is correct, the non-redundant quasi-realist must not only believe in hyper-external reasons but also defend our right to believe in them given the quasi-realist conception of the point of evaluative practice. This is none too easy a pair of things to do.<sup>33</sup>

<sup>33</sup> Special thanks to Paul Boghossian, Susan Hurley, Mark Kalderon, David Lewis, Denis Robinson and Michael Smith. A version of most of this paper was given as part of a three-lecture series at New York University in the fall of 1987. I have been much helped by that occasion. Thanks also the members of the Society for Ethical and Legal Philosophy (SELF).