# 3

# Naturalizing Norms

> We are all of us born in moral stupidity, taking the world as an udder to feed our supreme selves: Dorothea had early begun to emerge from that stupidity, but yet it had been easier to her to imagine how she would devote herself to Mr Casaubon, and become wise and strong in his strength and wisdom, than to conceive with that distinctness which is no longer reflection but feeling—an idea wrought back to the directness of sense, like the solidity of objects—that he had an equivalent centre of self, whence the lights and shadows must always fall with a certain difference.
>
> George Eliot, *Middlemarch*, ch. XXI

## 1. IN THE BEGINNING WAS THE DEED

The natural world is the world revealed by the senses, and described by the natural sciences: physics, chemistry, and notably biology, including evolutionary theory. However we think of it, ethics seems to fit badly into that world. Neither the senses nor the sciences seem to be good detectors of obligations, duties, or the order of value of things. As everyone knows, nature is heartless; the universe runs as much in accordance with its own laws when it brings suffering and ruin, as on the occasions when it brings peace and prosperity. Human beings too run as much in accordance with their own mixed and fallen natures when they do each. Iago is just as natural as Mother Theresa, and on a head-count perhaps more so. It may once have been a consolation, but it is so no longer, to think that the order of the universe is an ethical order. It is not, and even if it were, we would have no access to what the order is.

To be a naturalist is to see human beings as frail complexes of perishable tissue, and so part of the natural order. It is thus to refuse unexplained appeals to mind or spirit, and unexplained appeals to knowledge of a Platonic order of Forms or Norms; it is above all to



refuse any appeal to a supernatural order. After that, the degrees of austerity that naturalism imposes can be variously interpreted: some philosophers are more relaxed than others about reconciling the world as we know it, 'the manifest image', with the world as science tells us it is, 'the scientific image'. But we nearly all want to be naturalists and we all want a theory of ethics. So the problem is one of finding room for ethics, or of placing ethics within the disenchanted, non-ethical order which we inhabit, and of which we are a part.

'Finding room' means understanding how we think ethically, and why it offends against nothing in the rest of our world-view for us to do so. It does not necessarily mean 'reducing' ethics to something else. Reductionism here, as elsewhere in philosophy, implies seeing one thing as if it were another. Fastidious philosophers are rightly suspicious of it: as Moore famously said, quoting Butler, 'everything is what it is and not another thing'.[1] Nevertheless, the reconciliation of the normative and the natural must be carried out somehow, so if we are not reductionists we must find some other strategy.

I said in the first chapter that ethics was more a matter of knowing how (to behave), or knowing whom (to defer to, or punish, or admire), or knowing when (to act, or withdraw), than a matter of knowing *that* something is the case. Ethical knowledge, unlike knowledge of physics or history, can be quite inarticulate, and a novelist can paint a subject's ethics without ever showing them saying anything ethical. Whereas you could not paint a subject's knowledge of physics or history without showing them saying things belonging to those disciplines. If ethics is not in fact inarticulate, this is because we need to discuss how to behave and whom to admire, and to pass on the solutions to such problems that we find. Ethical sentences are the focus of these transactions.

The theory I want to defend is one that gives a story about the way in which ethical thought functions. Valuing something, it says, is not to be understood as *describing* it in certain terms, any more than hoping for or desiring something are describing it in particular terms. Rather, the state of mind of one who values something is distinctive, but nevertheless it is itself a natural, and naturally describable, state. Once we find ethics here, we understand the essential phenomenon, which is that of people valuing things. When they value things, they express themselves in terms of what is good, bad, obligatory, right, justifiable, and so on. When they wonder what to value, they express

[1] G. E. Moore, *Principia Ethica*, epigraph.

themselves as not being sure what is good or justifiable; when they achieve a certain kind of confidence, they say they know what to value, or what is valuable. The ethical proposition gets its identity as a focus for practical thought, as people communicate their certainties, insistences, and doubts about what to value.

This strategy—that of expressivism—leaves ethical properties and propositions alone with their own specific identities. They are the counters in our transactions with our values, just as a piece of money is a counter in financial transactions. To understand the value of a piece of money it is no good staring at it. It is necessary to understand the processes of human economic behaviour. You need to approach the token not with a microscope and a scalpel, but with an eye for large patterns of human interactions. Similarly, to understand the ethical proposition, it is no good looking for a 'concept' or 'truth condition'. We need the same eye for whole processes of human action and interaction. We need synthesis, not analysis.

So the expressivist thinks we can say interestingly what is involved for a subject *S* to think that *X* is good. It is for *S* to value it, and this can be explained in natural terms. Nature itself may be heartless and free of desires, but amongst the creatures it has thrown up are some which are not heartless, and not free of desires. We understand our values by understanding ourselves as valuing, and this we can do. If you go on to ask this strategist what it is for something to *be* good, the response is that this is not the subject of this theoretical concern—that is, not the subject of concern for those of us who, while naturalists, want a theory of ethics. Either the question illegitimately insists that trying to analyse the ethical proposition is the only possible strategy, which is not true. Or it must be heard in an ethical tone of voice. To answer it then would be to go inside the domain of ethics, and start expressing our standards. In this sense we may discuss whether promoting human flourishing, or manifesting respect for nature, or for liberty and equality, are good. But this kind of discussion is not furthering the project of explaining ethics in natural terms. It is taking ethical thought for granted, and trying to express and systematize our actual values.

Expressivism denies that when we assert values, we talk about our own states of mind, in actual or potential circumstances. It says that we *voice* our states of mind, but denies that we thereby describe them.[2] Similarly, if we are sincere when we say that 'the time is midnight' we voice our belief, but we do not describe ourselves as having a belief.

[2] Huw Price, 'Truth and the Nature of Assertion', *Mind*, 96 (1987), 202–20.



Our having the belief is not what makes it true that the time is midnight. It is only what makes us sincere when we say it.

Wittgenstein said, after Goethe, that in the beginning was the deed: *im Anfang war die Tat*. Words are themselves deeds ('words can be hard to say').[3] That is, it is only through understanding the *activities* associated with particular linguistic transactions that we understand the words used in conducting them. Amongst the activities involved in ethics are these: valuing, grading, forbidding, permitting, forming resolves, backing off, communicating emotion such as anger or resentment, embarrassment or shame, voicing attitudes such as admiration, or disdain or contempt, or even disgust, querying conduct, pressing, attack, warding it off. When I say that these are involved in ethics, I mean what I have already adverted to, that by describing the contours of a character in terms of doings like these, a narrator can tell us all that is important about the character's ethics, regardless of the words said.

Should all these activities be herded together as 'expressing ethical beliefs'? It is hard to see how that could be useful to do so. It would be labelling at a level of abstraction that makes the interesting detail invisible. A philosopher might carelessly regard this as harmless: perhaps he sees it as simply revealing the 'depth grammar' or 'logical form' behind the rather ragged surface of linguistic behaviour. But it is not harmless. When we voice our ethics we have a distinct conversational dynamics. People are badgered. Reproaches are made and rejected. Prescriptions are issued and enforced. Resentments arise and are soothed. Emotions are tugged. The smooth clothing of statements proposed as true or denied as false disguises the living body beneath. The expressivist task is to reveal that clothing for what it is—but that is not to say that we should always try to do without it.

## 2. PRELUDE: NORMS AND FUNCTIONS

The evaluative proposition is the focus for practical discussion: what to do, what to admire, whom to badger, when to repent, and so on. But although Wittgenstein may have been right that in the beginning was the deed, it is still true that deeds need interpretation. We do not know what someone is doing until we can see their mere movements as expressions of purpose and intention, conducted in the light of beliefs about their situation, desires, and emotions. So before tackling

[3] Ludwig Wittgenstein, *Philosophical Investigations*, trans. G. E. M. Anscombe (Oxford: Blackwell, 1953), § 546.

the evaluative proposition head-on, I wish to consider for a while the theory of interpretation in general. What are states of mind?

States of mind are natural states. They are extremely hard to define. There is one well-known reason for this: the famous 'holism of the mental'. This view can best be understood by contrast with the view that it superseded, which was the view often called logical behaviourism, particularly associated with Gilbert Ryle.[4] Logical behaviourism proposed that each state of mind of an individual, such as their desiring this or believing that, could be defined in terms of the characteristic behaviour that would be expressive of that state of mind. The objection on the grounds of holism is that there simply never exists a characteristic piece of behaviour of the kind required. How people behave in the light of their beliefs depends upon what they desire. But how they behave in the light of their desires depends upon what they believe. If I believe that it is a fine day, this may prompt me to go for a walk, but only if I like that kind of exercise, and if nothing else interferes. If I fear ultra-violet light, it might make me stay indoors. It might make me say 'It is a fine day', but only if I want to tell the truth; others might be prompted to deny that it is a fine day, depending on what stratagems they have in hand. And so on. On this view a person's entire mentality forms a kind of web or field or force in which no single element has its own self-standing connection with action. Different beliefs and desires (and perhaps other states, such as emotions, attitudes, wishes, fantasies, fears, and of course values) come together to issue in action. But the contribution of any one of them will vary according to what else is in the mix, and therefore resists definition in terms of behaviour.

Holism means that we must beware of over-simple connections between values and action. But, in spite of holism, values, along with every other mental state, are eventually read back *en bloc* from peoples' doings. Values, like desires, are manifested in behaviour, even if they are only manifested as part of a crowd of other mental states, some of which, sometimes, do a good job of obscuring them.

Philosophers of mind have learned to live with the holism of the mental. Few deny the phenomenon, but few think of it as posing an

[4] The association with Ryle is loose. Ryle himself was at best a half-hearted reductionist in the philosophy of mind. First, his conception of a 'multi-track' disposition shows a realistic appreciation of the holism of the mental, and the diversity of ways in which a mental state may issue in actual behaviour, depending upon what other things the subject believes or desires. Secondly, Ryle is quite happy to use mental vocabulary on the right-hand side of his equations. There is no implication of reduction to behaviour as it is witnessed by the camera or Martian.



insuperable threat to any kind of naturalism. It shows that it was wrong to go for a simple item-by-item analysis of mental states in terms of overt behaviour. But it does not show that it is impossible to isolate the function of a particular belief or desire in the rest of a cognitive economy. And it emphatically does not show that the whole system is somehow unnatural, or magical, or resistant to natural identification.

But there is another aspect, which we can call the pervasiveness of normativity. The idea is this. Holism does not seriously threaten our naturalistic hopes in the philosophy of mind. It only suggests that the mind is like a complicated causal field in which many factors combine to influence any outcome. A belief or desire can be defined by its effect over all such fields: what it would do if found in combination with one or the other mix of other beliefs and desires. The pervasiveness of normativity denies that it is just like that. What it suggests instead is that when we are in the domain of the mental we are in an order of *rationality* rather than an order of *causality*. In Sellars's famous phrase, we are in the 'logical space of reasons'.[5] In other words, we can say what a person *ought* to do or what it *makes sense* for them to do if they have such-and-such a belief, coupled with such-and-such other beliefs and desires. But we cannot directly, on this account, say what they *will* do. Mental states are located not by their place in a causal structure, but by their place in a rational structure. And 'rational' here means normative: it tells us how it would make sense for a person to factor a belief or desire into a pre-existent matrix of mental states. A desire for an apple is something that makes sense of an agent's action of picking up the apple, given of course a normal human background of other beliefs and desires. But it equally makes sense of an agent's avoiding the apple, given that he has an overwhelming religious desire to mortify the flesh.

This way of looking at it is encouraged by much writing about the situation of the 'radical interpreter'. This is someone thought of as facing the problem of identifying a subject's beliefs, desires, and other mental states on the basis of their linguistic and other behaviour, but without prior acquaintance with the language they speak. It is supposed that investigating the *epistemology* of this enterprise is the best way of understanding the *ontology* or nature of mental states in general (this is the doctrine called interpretationism in the philosophy of

[5] W. F. Sellars, 'Empiricism and the Philosophy of Mind', reprinted in *Science, Perception and Reality* (London: Routledge & Kegan Paul, 1963), 169. Sellars actually only makes the claim about imputations of knowledge, but it is natural to generalize it.

mind, most notably associated with Donald Davidson and Daniel Dennett).[6] It is then added that the epistemology is best seen as normative in its essence. It is a matter of 'rationalizing' the subjects, hypothesizing that they believe what they ought to believe, and desire what they ought to desire, or at least what it makes sense for them to desire. Such a principle is known in various forms as a principle of charity or principle of humanity. Given this, mental states turn out to get their identity from a network of normative considerations.

There are two more buttresses to this view that I shall mention. One is that any exercise of judgement requires a subject who is not only disposed to apply or withhold terms, but who regards it as *correct* to apply or withhold terms. Speaking meaningfully is more than regularly producing noise. What animates words or gives them their life is the way in which their application is correct or incorrect, and this means that they are subject to evaluation. Obviously in human life this evaluation is a public matter and involves a community of speakers, although it is controversial whether this is necessarily so. But even if one lonely speaker, a congenital Robinson Crusoe, is a possibility, if he or she uses words then he or she must have established norms for their correct application. A second buttress to the view is the plausibility of a theory of interpretation also allied to that of Dennett and Davidson. On Collingwood's account, to understand your doings and sayings I must see them as the things *to have* done and said in the light of your grasp of the situation.[7] And that means making your words and sayings my own, in the sense of thinking the situation through 'in your shoes' or from your point of view, and then realizing that you did what was to be done, or that you said what was to be said. That process of rationalization also requires evaluation, in this case my own evaluation of sensible actions and the appropriate applicability of terms in your situation. It does not prevent me from discounting for things, say, that I know and you do not. But it does require me taking my own thinking and reasoning skills into your situation. In so far as I cannot see myself saying or doing what you did on any account of the matter, I do not understand you.

It is essential to realize that, in the view of writers in this tradition, the principle of charity is not here an optimistic and optional assumption about how well we are doing. It is a 'constitutive' rule, or a principle that governs the very essence of mental states. It is not open to empirical rebuttal: it is a tautology, or principle that defines its subject-matter. Writing it in a form on which we can focus, we have an a priori principle of interpretation (API):

> (API) It is analytic that creatures with beliefs, desires, and other states of mind, behave in ways that (best) make sense (and not in ways that make no sense), given those states of mind.

The idea behind calling this a constitutive rule is that it tells us *what it is* to have beliefs and other states of mind. According to API, then, it is analytic that creatures *conform* to the normative or rational order. A creature which appears not to do so is either a creature that we have misinterpreted, or a creature that has no mental states, but merely exhibits movements.

With API in place, it turns out not to be a simple empirical matter to describe a creature as believing this and desiring that. It turns out that issues of attitude are involved. To interpret someone as believing that the glass contains benzene and to loathe the idea of drinking benzene is partly to be poised to inveigh against them if they drink what is in the glass. But this is just what is involved in locating mental states in a normative space, rather than in purely causal terms.

The normative dimension to interpretation might seem to be consistent with a view of the normative which is non-naturalistic. That is, the existence of 'oughts' of rationality could be a real, Platonic mystery, for all this approach tells us. They could obtain in virtue of highly mysterious, *sui generis* facts about the rational order: facts which bear only a strange relationship to the natural order, and whose own credentials and authority remain shrouded in obscurity. If we think of it like this, then the expressivist strategy for naturalizing values runs into trouble. For it pins its faith on the possibility of a natural account of the state of valuing something. If no such account is forthcoming, because valuing, like other mental states, is identifiable only through the existence of a network of norms whose own nature is entirely mysterious, then the promise collapses.

But this is over-pessimistic. Suppose a principle or 'platitude' governing how to attribute a mental state such as fear of dogs to someone contains clauses such as this:

> $X$ fears dogs $\equiv$ $X$ is in a state which, in conjunction with the belief that a dog is about, means that (other things being equal) it makes sense for $X$ to flee; in conjunction with the belief that the

[6] For Davidson, see the essays collected in *Essays on Actions and Events* (Oxford: Oxford University Press, 1980). For Dennett, see *The Intentional Stance* (Cambridge, Mass.: MIT Press, 1987).

[7] R. G. Collingwood, *The Idea of History* (Oxford: Oxford University Press, 1946).

dog's owner is present, means that it makes sense for X to ask him to hold the dog, etc.

Then the question is how to take the references to 'making sense' in these clauses. They seem to be statements about what you would expect from an agent. It would surprise you, and lead you to scramble for another interpretation of X, if X does not behave in ways that make sense. Similarly we might say that the circuit is wired up if it is in a state such that, when the key is turned, there ought to be a spark. But the 'ought' here speaks not of duties and values, but just about what you would expect. If there is no spark, then something has gone wrong, but this too means simply loss of expected or intended function. The ignition creates the current; that is what it is for. The fear prompts the flight; that is what it is for.

To focus our thoughts, let us return to the apparent rival to the normative way of thinking about the mental. The functionalist rival at first either rejects the pervasiveness of normativity, or defuses it in the way suggested by the example of the circuit. It simply cites what we think mental states *do*, not what we think they ought to do. According to functionalism, any state of mind is located in essence by a trio of causal relations the state may have. It is caused by some things; it causes other mental states; and in conjunction with other beliefs, desires, and values it has various effects on action. The state is located by its typical causes, its mental relations, and behavioural effects. Saying that if you fear dogs, then when you come upon one you ought to flee, is just saying that this is the kind of behaviour you would expect of people who fear dogs. There is no implication that if you do not flee, you are somehow irrational. You are only to some extent doing something unexpected, to be explained with further investigation, and even indicative that your fear of dogs is not all that real.

Functionalism identifies mental states by their place in causal networks, not by their place in systems whose principles of construction are normative. We can 'solve for' fear of dogs at the same time as we solve for other elements of such a network. But in all of this there is nothing specifically normative involved. There are just dispositions to actions, and solutions for the elements in the causal field that give rise to action. States of mind are just there, identifiable in the psychologies that make people act as they do. Of course, the functionalist can add to the story some of the ingredients that impress those who put normativity in a central place. We represent the world to ourselves. If we represent it wrongly, plans and projects based on the misrepresentation are apt to fail. If too many fail, evolution sweeps us away. If our



desires do not bear some relationship to our needs, the same happens. We ourselves deploy norms in these areas: we decry badly formed belief as misrepresentation of the world; we decry non-adaptive desires as disagreeable or useless to those who have them and those around them. There is nothing unnatural in our valuing accurate belief and healthy desire, once valuing is itself registered as one natural state among others.

Is there any opposition between the first, norm-centred approach to rationality, and the second, cause-centred approach? The first is, apparently, evaluative where the second is, apparently, empirical or causal. But this opposition seems to be markedly softened when we take into account API, the 'constitutive' role of the principles of rationality. For API means—this is something of a mouthful—that it is analytic that (typically) creatures exhibit a causal structure that is *isomorphic* with the rational structure that the normative approach prizes. For example, suppose it makes sense for Fred to drink a glass of benzene because Fred wants to drink a glass of gin, and has been plausibly but misleadingly told that the glass is full of gin. And suppose it would make no sense at all for Fred to drink what is in the glass if the deception is revealed. Given API, Fred's desire for gin will cause him to drink the liquid in the first scenario, but will not cause him to drink it once the deception is revealed. If Fred did drink the liquid after the deception comes to light, we have to search again for interfering factors: perhaps this makes sense because Fred wants to see what happens, or wants to exhibit a reckless machismo. Given such wants, it makes sense for Fred to drink the benzene, and if Fred does so it will be one of these wants that explains why he does so. In such a way, it is not only true, but a matter of definition or analytic, that Fred's desires and beliefs form a causal structure that mirrors the structures of what it makes sense for him to do. If they don't, then we have to reinterpret him.

How can API be true? How can it be a matter of trivial, analytic or definitional truth that people conform to a certain normative order? It would seem to be contingent, and even on the face of it optimistic, to suppose that by and large they do what they ought to do, in the domain of rationality just as much as in that of morality. But we can answer the question like this. Consider the example of game-playing. Here, too, there is a definitive normative order: a game is defined by its rules. There is a limit to the extent to which people can fail to conform to the rules. A rule may be broken now and again, but systematic and acknowledged breaking of the rules becomes not that, but a change of the game. Yet it is largely an empirical matter which game

people are playing. Their behaviour tells us which patterns they *do* conform to and that in turn tells us which game they are playing: that is, which are the rules to which they *ought* to conform. Roughly: to play football is to behave in any of the large variety of ways that are licensed by the rules of football. It includes avoiding those actions that are prohibited by the same rules. It is also, in self-conscious beings, to have behaviour guided by explicit or implicit knowledge of what the rules are. There cannot be a total mismatch between what people in fact do on the playing field and what they ought to do. Similarly, to have the desire for *F* is to be disposed to behave in any of the variety of ways that are made sense of by having the desire for *F*. To have the value *V* is to be disposed to behave in those ways that are made sense of by having that value. There is, in other words, an assimilation of the normative and the causal order. We know what a desire is by knowing what it would make sense to do in the light of having the desire; but then we know whether someone has the desire by seeing if this light is one that makes good sense of what they do. API can be true because desires and beliefs are defined by what it is that they make sense of. But they are attributed by what they make people do, under the rubric that people do what makes sense to them. If we try to suggest cases where creatures (systematically) do what makes no sense to them, we end up denying the application of beliefs and desires at all. We are left only with behaviour, bodily movements that are not actions. This being so, the 'rational order' talked of in the theory of interpretation cannot be any very mysterious order of normative structures. For, if it were, it would inevitably be contingent whether people manage on the whole to conform to it—and that conflicts with API.

The theory of mind that is suggested by these considerations is twofold. On the one hand, the concept of a belief or desire, or any other state of mind, is identified using normative terms. These are defined in terms of what it makes sense to do in the light of them, given other states similarly defined. On the other hand, their presence in any subject is identified empirically, in terms of the causal structures visible in the actions the subject performs, and those she would perform in other circumstances. Any apparent mismatch is averted via the analytical principle API.

Where computation fails in the rational order, rules for prediction fail in the natural order. Computation fails because we may not be sure in advance how new cases may strike us, or strike others. This means that we recognize the *permeable* nature of the mind: new things catch our attention. New experiences, new angles of vision, new comparisons, new wishes, fantasies, memories, moods, emotions occur, and each of these can influence what it seems to us to make best sense to do. Revisiting the factors in a deliberation is like revisiting old friends. Much of what happens is predictable, but surprises and revaluations do occur, and are always possible. But the causal order shows exactly the same permeability, for exactly the same reason. When new things catch our attention, or we become subject to new moods or memories, they influence what we think it makes sense for us to do, but, equally, they influence what we do in fact do. They act as new causal factors. So uncodifiability in the rational domain is matched by causal complexity in the natural domain, and indeed, given API, this has to be so.



We draw some morals from the theory of interpretation when we look at preferences, utilities, and choices in Chapters 5 and 6. We shall find that the theory of expected utility conforms to this model, imposing a parallel logical or normative structure on agents whose behaviour makes them eligible for interpretation as agents in the first place.

## 3. STATES OF MIND: SATAN AND OTHELLO

Let us now concentrate on what we regard as a person's values. Expressivism requires a naturalistic story of the state of mind of valuing something. We then go on from that to give an account of the procedures of valuation that we adopt; the modes of expression that are appropriate; and finally the logic and theory of meaning of our typical expression of values. We have already seen that we locate a person's values in the light of a number of manifestations: what they say, what they do, what they regret in themselves, what they encourage in others, what they forbid or what they insist upon. Sometimes these elements pull together, and we have no doubt what someone values. Sometimes they do not harmonize all that well, and suggest various interpretations. Someone may sincerely believe that something is best, but not do it. There may be states of loathing of ethics, or of desire to be bad, or of maliciousness or waywardness, as well as despair or lethargy (what Aquinas called *accidie*), all of which can interfere with a simple attempt to read anyone's values straight back from their choices. Nobody lives up to their better selves all the time; some people only do it very little of the time.

There are interestingly different, although related, ways of interpreting a lot of such cases. If a person fails to live up to their professed values enough of the time, we start to doubt whether the professions

are sincere, or, if they are at least sincere, we may wonder about self-deception. One class of cases is the simple 'inverted commas' type of case, where an agent pays lip-service to a value that they do not really hold, either through hypocrisy or self-deception. This is no problem: a person says that an action other than the one he intends is the 'good' action, and means only that it is what the others call 'good' (there is a specific sneering intonation that is typically used when we so speak). Another very different class of case is where something is given an evaluative label through *inertia*, even after the usual connection between valuation and motivation has been severed. For example, we deem a wine good in light of the pleasure we take in tasting it. This is the typical or basic case. But if, for some reason, we have lost our taste for wine and take no pleasure in tasting any, we could still call some wine good if we know that it once merited the label. This has an obvious point: it serves the public function of grading the wine, or encouraging others to try for the pleasure. A third class of case I would like to distinguish is different, although related. Here evaluative words have also 'gone dead' and retain a use only to specify the class of things meeting the standards that apply. Consider this conversation from Jane Austen's *Emma*:

> 'Mr. Dixon, you say, is not, strictly speaking, handsome.'
>
> 'Handsome! Oh! no—far from it—certainly plain. I told you he was plain.'
>
> 'My Dear, you said that Miss Campbell would not allow him to be plain, and that you yourself—'
>
> 'Oh! as for me, my judgment is worth nothing. Where I have a regard, I always think a person well-looking. But I have what I believed the general opinion, when I called him plain.'

Here Jane Fairfax, the second speaker, does not put inverted commas around 'the general opinion', yet the word 'plain' is applied in accordance with it, rather than as an expression of her own judgement. The use is in a certain sense deferential to normal opinion. But, of course, the valuation still lies in the background. It provides the reason why Mr Dixon is called plain even if, in the mouth of Jane Fairfax, it expresses no aesthetic attitude of her own. Jane is part of the social practice of rating people on their looks. If she is spectacularly out of line with the others, she will be criticized by them for misleading them, and this is what she seeks to avoid in the conversation.

Other cases are more interesting, because a person's own values are involved, yet there is only a shaky or perverted connection with motivation. On occasion, clearly, a person can act, knowingly and intentionally, against her values: she may have desires that overcome



her scruples, or just knowingly succumb to temptation. Doing wrong even has its own allure. But if the values are really there, we will expect them to manifest themselves somewhere else: in regret or remorse, or in guilt or shame. Yet even that may be too simple, and some philosophers ('externalists') have suggested that there can be agents in whom a pure cognizance of ethics has no practical effect: they know what is good to do, and simply do not care. Perhaps it is only good people who care to do what they know to be good. Perhaps there are bad people who know what it is good to do and then deliberately direct their wills the other way. Thus in Book IV of *Paradise Lost* Satan describes his motivation for bringing about the Fall with the chilling resolution 'Evil be thou my Good' (l. 110). For Satan, the judgement that something is evil acts as an attraction. And the fact that this possibility makes sense casts some doubt on the very close identification I have been urging between ethics and practical, motivating states of mind. If externalists are right, perhaps we have to see ethics more in terms of awareness of fact, with it then being up to us whether we care about the perceived fact one way or another. Philosophers resisting this ('internalists') have to say how they interpret the persistent, careless person who sees what is good, and doesn't care, or who, like Satan, sees what is good, but goes the other way.

My own judgement on this debate is that externalists can win individual battles. They can certainly point to possible psychologies about which the right thing to say is that the agent knows what it is good or right to do, and then deliberately and knowingly does something else. And they can point to psychologies like that of Satan, in which it can become a reason for doing something precisely that it is known to be evil. But internalists win the war for all that, in the sense that these cases are necessarily parasitic, and what they are parasitic upon is a background connection between ethics and motivation. They are cases in which things are out of joint, but the fact of a joint being out presupposes a normal or typical state in which it is not out.

To understand this, it is useful to think of another analogy. Consider the complex of dispositions involved in being in love with someone. This typically includes taking pleasure in their company, wanting above all to be with them, wanting to give them pleasure and take pleasure from them, and so on. Nevertheless there are cases in which one person is in love with another, but wants not to see them (it would be too painful), or even wants to hurt them (jealous revenge). These cases are necessarily parasitic upon the normal in the sense that they require a background of the normal dispositions, which have then been wrenched out of order, giving rise to jealousy

or the desire for revenge: love coexisting with hatred. But it would be absurd to conclude that being in love with someone is therefore a purely cognitive state, having no necessary connection with emotion or attitude.

We can approach the issue by another ethical illustration, here a public act rather than a private attitude. What is it to forbid something? To issue an injunction against it—but what is that? We might talk of communicating an intention to invoke sanctions or to become in one way or another set against anyone who disobeys the injunction. Surely 'forbidding' inhabits that neighbourhood. But can't you forbid someone from doing something, while all the time intending to forgive them if they do it? The case skates perilously close to pretence, or play-acting. But we might allow it: parents, for example, seem to tell their children what not to do, but without any apparent intention of doing much about it when they are disobeyed. If we allow it, we should say that the case exists, but is parasitic on a more robust social connection between forbidding and the disposition to sanction. The link is put out of joint by half-hearted parents, but it exists as the background to their activities, and is necessary to make those activities possible. Once more, it would be wrong to conclude that there is no necessary connection between an act of forbidding and an intention to invoke a sanction. That has to be the typical case that any others exploit.

Coming back to the individual psychology, Milton's Satan marvelously illustrates the point. Before saying 'Evil be thou my Good' he makes the terrible renouncement:

> So farewell Hope, and with Hope, farewell Fear
> Farewell Remorse: all Good to me is lost . . .
> (*Paradise Lost* IV. 108–9)

and Milton tells us:

> Thus while he spake, each passion dimmed his face
> Thrice chang'd with pale, ire, envie and despair,
> Which marred his borrowed visage, and betrayed
> Him counterfeit, if any eye beheld;
> For heav'nly minds from such distempers foul
> Are ever clear . . .
> (IV. 114–19)

Satan is racked with all the 'foul distempers' going. His position is exactly analogous to that of the jealous or despairing lover who wants to harm the beloved. It is his perpetual curse that he remembers how



things could have been otherwise (he and God, representing the Good, could still have been together, as they once were); Milton's whole poem describes how his resolution to do evil is a response to this nightmarish predicament. It is a reaction against an acknowledged, internalized, active set of concerns, which align with those of the good. This is why his predicament is one of rebellion not only against the good, but against part of himself. In other words, this is not a psychology in which 'knowledge of what is good' is emotionally or practically inert. It is a psychology driven by reaction *against* the emotional and practical force that knowledge exerts. Such a psychology represents an ultimate conflict within the agent, and perhaps is as likely to issue in suicide as murder. The point is made in the myth: Satan's knowledge of what is good is established by his being a *fallen* angel, in whom the appropriate attitudes and emotions were once in place.[8] Similarly, in order to be jealously murderous, a lover must once have loved in more normal ways, and been dislocated from them.

Is it a conceptual truth, or a truth of empirical psychology, that the subject must originally be one way, and only later get out of joint, for these interpretations to earn their keep?[9] If the latter, we might talk of 'projection' or 'dislocation' simply as psychological mechanisms, useful for explaining how a good person becomes attracted to the bad or how a lover becomes murderous. If the former, we would deny that there are even possible cases in which the love of the bad comes about except through a previous love of the good going out of joint. My own opinion inclines to this, although it is not essential to the main message of expressivism. The main message can allow that in principle an individual may love the bad without ever having loved the good, provided enough social context, in which motivations are aligned with the good, is provided.

In Shakespeare's play, Othello still loves Desdemona as he smothers her. One might prefer to say that his love has turned to hate. But I think this would be simplistic, or even just unkind (for these verdicts will carry moral implications).

During the crisis, Othello is never simply somebody who hates Desdemona, and to say that would not explain the eventual size of his remorse when he knows what he has done. His love is far from

[8] The phenomenon is superbly treated in Salman Rushdie's *The Satanic Verses*, 424–7.

[9] I am grateful to Margaret Walker for forcing me to confront this question.

extinguished, but only temporarily overcome. Even when he is accusing her, Desdemona is

> There where I have garnered up my heart,
> Where either I must live or bear no life,
> The fountain from which my current runs
> Or else dries up—
>
> IV. 1. 59–62

And after the murder:

> Had she been true,
> If heaven would make me such another world
> Of one entire and perfect chrysolite
> I'd not have sold her for it.
>
> V. 2. 150–3

The problems with Othello are his fears and his conception of honour, not his attachment to Desdemona. But we look further in Chapter 5 at the sinister way in which egoistic motives might underlie even the most fervent and sincere attachments.

In any fully realized case of love behaving as hatred does, or of evil being someone's good, we can provide more or less plausible empirical speculations about how the joint gets to be out. It is here that we talk of displacement and egoism and pride and immaturity, or whatever strikes us as the best history of the conflicts within the breast.

Of course, there are many less extreme examples of dislocation between ethics and action. I might recognize that it is my duty to do something, and resent that fact because I resent having been put in that position. I still have to be the victim of a foul distemper for this to happen.[10] But I might just feel like being mischievous, or even malicious for once (this can be *delicious*). If I am very dutiful normally, but the others are having more fun, I might want to join them. But, again, these cases only exist against either a *psychological* background of motivation by what is perceived as duty, or a *social* background of insistence upon duty as a practical constraint. That is, even if we try to stretch the case away from the Satan/Othello model by sketching an agent who has absolutely no conflict, but views duties with the utmost unconcern, we will find that he exists only against a backdrop in which talk of duty does express concern.

If there is nothing but settled, cold unconcern from an agent for

[10] Hume describes the nice case of an historical situation in which because of too much prating about ethics, a 'peevish delicacy' led people to belittle its claims: *Enquiry Concerning the Principles of Morals*, VI. 1, p. 242.



what he verbally acknowledges as his duty, then of course we do begin to talk of mere lip-service. The agent is using evaluative vocabulary in a parasitic way, as mere labelling for what other people regard as good. He may be like Jane Fairfax, or the wine-taster who has lost pleasure in wine, using the word as a label for the kind of wine in which he would expect others to take pleasure. But we are, rightly, not quick to reach for these interpretations in ethics. Partly this is because, if the agent has a normal history, we would expect internalized values still to operate at some level. There will be incipient conflict. Partly it is an exercise of a general interpretive strategy, which is to hold people to the literal meanings of their words. Unless agents deliberately indicate the inverted commas we take them literally, as indeed Jane Fairfax was taken in the conversation quoted above.

Doing what you know to be bad is bad. We might describe it as irrational, and since Plato many philosophers have done so. But that is not, as it stands, a very interesting thing to say, for it is not at all obvious what further or different specific kind of charge it makes. Satan's pride got him where he is, but his pride is not inexplicable or unintelligible. It is not as if he insists that $7 + 5 = 13$, or that he is a teapot. In fact, Milton's poem is so great because Satan is not only intelligible, but not wholly unadmirable. His reaction to God's suffocating superiority makes sense. If a loving but strict parent lays down all kinds of constraints on a child's conduct, we understand the child being naughty and we might admire her for it, even if, as she is being naughty, she is in conflict with herself and as a result unhappy. In fact, parents are often more or less quietly proud of naughtiness in their children. Rebelliousness takes courage. Similarly, jealousy may be a very bad emotion to feel, but it is not clear what is meant by saying that it is irrational to feel it. It makes sense, in human terms, as a response to fear of loss or betrayal, even if heavenly minds are clear of it. Satan is not irrational, but bad. And if that is true of Satan, it is still more so of the more mundane cases of human weakness: we are depressed, we let ourselves be led, we succumb to temptation, we even make little attempts to break the moral mould. We are not irrational, but weak. Or, if 'rational' implies boundless strength of mind in pursuit of the good, say that we are irrational. But we are perfectly intelligible, and may deserve sympathy, and if our characters require improvement, it is not usually in the direction of better calculation or logic.

So externalists can have individual cases, but internalism wins the war. Ethics remains essentially practical, a matter of attitude, disposition, and emotion. When a psychology contains other elements, the

way these issue in action becomes surprising, or even in a way contradictory, in the way that Othello's final expression of his love of Desdemona contradicts that emotion. And that is just the way it is with the relationship between attitude and action, in beings as complex as we are.

How, then, do values relate to other practical states, and notably to desires? Do they constitute a special *sui generis* motivational state, one of 'normative governance' rather than normal desire? Gibbard talks of a state 'identified by its place in a syndrome of tendencies towards action and avowal—a syndrome produced by the language-infused system of coordination peculiar to human beings'.[11] I think this is right, although I shall later caution against supposing that the motivational system in question is special or extraordinary. And I shall return to some of the detail of coordination later. But I doubt if anything more direct would be right. Philosophers have often suggested equating an agent's values not simply with what she desires, but with what she desires to desire, or with her highest-order desire, thought of as the desires she 'identifies with', or has no tendency to wish away.[12] This, too, is surely along the right track, although it does not work well with Satan. Satan does not at any level desire to desire as the good angels do, happy indulging in monotonous worship, even although he knows that in rebelling against that status he rebelled against being good. He feels the most intense pressure towards repentance, yet he does not desire to repent, and neither does he desire to desire it. This is partly because he is well aware that if he were to repent and become good again, the whole process would start all over again (IV. 93–100). So Satan's genuine present sense of evil and of the good he has lost does not coincide with present *desires* at any level, although it may align with regret and wishes that things might have been otherwise.

Our values also do not seem quite the same as our highest-order desires for another reason. It seems natural to interpret some agents as regretting having the values they do: a person of unswerving integrity, for instance, might at some level wish that he was a little more relaxed, if those who cut corners or bend the rules just a little seem to be having more fun or getting on better in the world.

[11] Allan Gibbard, *Wise Choices, Apt Feelings*, 75.

[12] The classic papers by Harry Frankfurt, particularly 'Identification and Wholeheartedness', are collected in his *The Importance of What we Care About* (Cambridge: Cambridge University Press, 1988). See also Gerald Gaus, *Value and Justification* (Cambridge: Cambridge University Press, 1990).



Obviously, however, if this wish starts to dominate his practical life, then his integrity has been compromised, and we might prefer to say that his values have changed, at least temporarily.

I may desire to eat ice cream, but be quite calm contemplating a future state in which I would not desire it. I need not resist processes that change my desires. In fact, I might welcome them. Whereas, when values are involved, we typically resist anything likely to destabilize them. Such processes would be regarded as undermining and threatening. So, if we imagine the general field of an agent's concerns, his or her values might be regarded as those concerns that he is also concerned to preserve: the ones by which he stands. He would contemplate losing them only with dismay. It is because a child or an animal has no such protective concerns that they are thought of as 'wanton'.[13]

To hold a value, then, is typically to have a relatively stable disposition to conduct practical life and practical discussion in a particular way: it is to be disposed or *set* in that way, and notably to be set against change in this respect. This way of being set is such as to align values and motivations. And, characteristically in philosophical psychology, although this may be the paradigm, typical case, there are also others. We are often not aware of what we value, or how much we value it. We may not be all that disposed to protect and cherish a concern, but, when things go wrong, realize how much it mattered. It may only be retrospectively that we discover that we have valued something. Here the disposition is not one of being set to foster a concern, but being set to feel pain when it is threatened or vanishes.

Because we are talking about the way an agent is set, there is a speed limit on change of values. Someone who one day professes one set of concerns and priorities, but another day a quite different set, is not someone who has many values, albeit for a short time each, but someone who has no real values at all. Over sufficient time, of course, a person's concerns and interests and priorities can change, although it is also worth noticing that values one thinks one has suppressed can be surprisingly resilient: habits of valuing, like habits of deference or old superstitions, tend to come back to haunt us. It is difficult to shake off their grip, and there is often something off-colour about the professions of a convert, just as there is about professions to be cured of different emotions.

What has been said here concentrating upon the notion of a value, which might typically get expression by saying that something is

[13] I am indebted to Valerie Tiberius for emphasizing this.

good, or that some character trait is a virtue, can be adapted quite easily to the vocabulary of right and duty, rights and obligations. Here we find states of mind further up the staircase of emotional ascent: ones that prime us to insistences and to hostility to others. But the connection with action is obvious in typical cases, and subject to the same kinds of interpretative subtleties when normal links are broken.

To sum up, then: to hold a value is to have a relatively fixed attitude to some aspect of things, an attitude with which one identifies in the sense of being set to resist change, or set to feel pain when concerns are not met. That fixed attitude typically issues in many dispositions, at various places on the staircase of emotional ascent I described in the first chapter. When things are not out of joint, values align with motivations. But we understand quite profound misalignments in terms of intelligible internal conflict. If the conflict is sufficient, we will not know what to say, just as we do not know when to say that love coincides with hate, or when it has been replaced by it. When things get out of joint the normal or typical expression of the attitude can be perverted; reactions set in, and under sufficient stress we have the person of weak will, who knowingly succumbs to temptation, or more interestingly the Satanic figure who is knowingly attracted to evil. To merit such an interpretation, however, and not to be put down merely as a hypocrite who pays lip-service to values he does not really hold, an agent must show a dynamic pattern of change and stress. There has to be a history, and an internal conflict. Not everyone who murders someone is murdering someone he loves, but occasionally some do, and sometimes we understand why.

## 4. THE ETHICAL PROPOSITION AND FREGE'S ABYSS

So what at last is said when we say that something is good or right? Following Moore, we do not expect to identify the content in other terms. We can now say, however, what is done when we say such things. We avow a practical state. 'Avowal' here means that we express this state, make it public, or communicate it.[14] We intend coordination with similar avowals or potential avowals from others, and this is the point of the communication. When this coordination is achieved, an intended direction is given to our joint practical lives and choices. Saying that something is good when we do not really value it is either deceiving others about our state, or is the result of self-deception. But because we have to accommodate the flexible, many-layered nature of our minds, we may sincerely say that something is good when we are not, unhappily, motivated to pursue it, provided one of the diagnoses sketched in the previous section applies.

If I permit smoking in my house, but you forbid it in yours, we do not necessarily disagree about anything. Similarly, some evaluations are happily relativized: the weather is good from the farmer's point of view, but bad from the tourist's point of view, or good in so far as it helps the crops, bad in so far as it spoils the holiday. But with much ethics there is no scope for this coexistence. If I am minded to permit smoking in our house, and my wife is minded to forbid it, we do disagree. Only one of these practical attitudes can be implemented, and I am for one, and she is for the other. When we discuss ethics with each other, we are typically talking about 'our' house, or in other words practical issues on which we want to coordinate, or have to coordinate. In that case difference of attitude means disagreement, just as surely as difference of belief does. If the case is like that of separate houses we can sometimes 'agree to differ', and drop the conversation. But sometimes, even if we do not have to coordinate our actions, we cannot agree to differ, for serious enough differences cannot be tolerated. I return to this when we discuss relativism, but meanwhile the point remains that the typical, default, position is that difference in attitude is treated as disagreement.

Ethical avowals, like decisions and verdicts, require grounds. If I grade one paper higher than another, I must be prepared to indicate some relevant difference between them. We acknowledge the need to point to something that grounds our judgement, in virtue of which one is better than the other. But, as we shall see in the next chapter, discussing Cornell realism, no complete theory of ethics can simply point to the grounding properties, and suppose that evaluations are given their meaning by their relationship to them. We need first to understand the evaluative stance.

Expressivism claims that the ethical proposition is something that we synthesize for a purpose. Its role is to act as a focus for practical thought. So what is it to believe that something is good, wonder whether it is good, to deny that it is good, to be undecided, or to know that it is good? In basic or typical cases:

[14] We are not prepared to avow all the norms that influence us. Gibbard makes the useful distinction here between accepting a norm, which includes preparedness to avow it, versus being in the grip of a norm, which means being subject to its sway, perhaps unwillingly or guiltily, as in the case of a not-quite-liberated racist or sexist, who knows how he ought to be, but does not quite measure up in his snap emotional reactions (*Wise Choices, Apt Feelings*, 58–61).

believing that X is good or right is roughly having an appropriately favourable valuation of X;

wondering whether X is good or right is wondering what to do/what to admire or value;

denying that X is good or right is rejecting a favourable attitude to X;

being undecided is not knowing what to do/what to admire, etc.;

being certain that X is good or right is having a settled attitude/rejecting the possibility that improvement could result in change;

knowing that X is good is knowing to choose X/admire X, etc.

Here the practical states on the right-hand side are voiced and discussed in terms of attitudes to the saying or thought on the left. This is what I mean by saying that the moral proposition is designed or invented or emerges naturally as the focus for our practical transactions. And in the previous section we established the natural credentials of the states on the right and their intricate connections with other attitudes and emotions.

The reason expressivism in ethics has to be correct is that if we supposed that belief, denial, and so on were simply discussions of a way the world is, we would still face the open question. Even if that belief were settled, there would still be issues of what importance to give it, what to do, and all the rest. For we have no conception of a 'truth condition' or fact of which mere apprehension by itself determines practical issues. For any fact, there is a question of what to do about it. But evaluative discussion just is discussion of what to do about things.

However, many writers have insisted on a 'Fregean abyss' separating expressions of attitude from expressions of belief. They point out that evaluative commitments are expressed in ordinary indicatives, and that this enables them to occur in indefinitely many 'indirect' contexts. We not only say that X is good, but that either X is good or Y is, or that if X is good such-and-such. Some think that this puts a weighty, or even insupportable, burden on expressivism. For all the expressivist has given us is an account of what is done when a moral sentence is put forward in an assertoric context: an attitude is voiced. What then happens when it is put forward in an indirect context, such as 'if X is good, then Y is good too' and no attitude to X or Y is voiced?

It is worth asking how other parts of language answer this problem. Suppose I say that the sentence 'Bears hibernate' expresses a belief. Well, it only does so when the sentence is put forward in an assertoric context. So what happens when it is put forward in an indirect context, such as 'If bears hibernate, they wake up hungry'? For here no belief in bears hibernating is expressed. The standard answer is to introduce a proposition or thought, regarded as a constant factor in both the assertoric and the indirect context. When we say bears hibernate, we express or assert the proposition, and represent ourselves as believing it; when we say 'If bears hibernate . . .' we introduce the proposition in a different way, conditionally, or as a supposition. Frege thought that in this second kind of context we refer to the thought that we assert in the assertoric context.

If this is allowed to solve the problem for ordinary beliefs, it might simply be taken over by the expressivist. In the Fregean story a 'proposition' or 'thought' is simply introduced as the common element between contexts: something capable of being believed but equally capable of being merely supposed or entertained. So why not say the same about an 'attitude'? It can be avowed, or it can be put forward without avowal, as a *topic* for discussion, or as an alternative. Just as we want to know the implications of a proposition or a thought, so we want to know the implications of attitude. What implies it, what is it right to hold if it is to be adopted?

If we want to know in other terms what is going on when we so put forward an attitude, we must look to the function of the indirect contexts in question. The key idea here is one of a functional structure of commitments that is isomorphic with or mirrored by the propositional structure that we use to express them. Thus someone may be what I called 'tied to a tree': in a state in which he or she can only endorse some combination of attitude and belief. Suppose I hold that either John is to blame, or he didn't do the deed. Then I am in a state in which *if* one side is closed off to me, I am to switch to the other—or withdraw the commitment. And this is what I express by saying 'Either John is to blame, or he didn't do the deed', or equally, 'If John did the deed, he is to blame'. By advancing disjunctions and conditionals we avow these more complex dispositional states. Taking advantage of the theory of interpretation sketched above, we can regard the state in question not just in functional terms, but also in normative terms. By using the disjunction I am presenting myself in a way that will deserve reproach and bewilderment if, without explanation, I go on to suppose both that John did the deed and is blameless. This makes no sense, unless I have changed my mind about something.

There has been some scepticism about whether this approach can deliver the mighty 'musts' of logic.[15] But we now see that it can do so perfectly well. Consider the example made famous by Geach, of inference according to the pattern of modus ponens. Someone saying each of '$p$' and 'If $p$ then $q$' has the premises of a modus ponens whose conclusion is $q$. He is logically committed to $q$, if he is committed to the premises. To put it another way, if anyone represented themselves as holding the combination of '$p$' and 'If $p$ then $q$' and 'not-$q$' we would not know what to make of them. Logical breakdown means failure of understanding. Is this result secured, on my approach, for an evaluative antecedent, $p$? Yes, because the person represents themselves as tied to a tree of possible combinations of belief and attitude, but at the same time represents themselves as holding a combination that the tree excludes. So what is given at one moment is taken away at the next, and we can make no intelligible interpretation of them.

We can put the point another way. A mental state, I have said, is identified by what it 'makes sense' to hold in combination with it. To avow a mental state is therefore partly to express acceptance of certain norms. To avow anything of the form 'If $p$ then $q$' is to commit oneself to the combination 'Either not-$p$, or $q$' and to be tied to that combination is to disavow the combination of $p$ with not-$q$. Holding both together is therefore unintelligible. Logic is our way of codifying and keeping track of intelligible combinations of commitment.

We might usefully compare the situation to that in the theory of probability. The basic psychological reality as we contemplate the chances, for instance in a horse race, may be one of vague differences of confidence, reflected as dispositions to bet (under idealized circumstances) at various prices, or within various ranges of prices. If we choose to voice these confidences, for instance by saying that Eclipse has no better than an evens chance, or that 100 to 12 sounds a fair price on Sunrise, then we enter a more structured normative space. We only make sense if the chances we assign to the different horses in the field obey well-known classical rules of probability, and those rules dictate inferences. For instance, in a two-horse race, if we think one horse has a better than evens chance, then we must infer that the other has not.

[15] Bob Hale, 'The Compleat Projectivist', *Philosophical Quarterly*, 36 (1986), 65–84. Related work includes G. F. Schueler, 'Modus Ponens and Moral Realism', *Ethics*, 98 (1988), 492–500; M. H. Brighouse, 'Blackburn's Projectivism—An Objection', *Philosophical Studies*, 59 (1990), 225; Nick Zangwill, 'Moral Modus Ponens', *Ratio*, 5 (1992), 177–93; Mark van Roojen, 'Expressivism and Irrationality', *Philosophical Review*, 105 (1996), 311–35.



Our dispositions to bet at different prices only make sense if they can be represented as beliefs in probabilities, satisfying those laws. Working in terms of 'belief in classical probabilities' does not, then, necessarily reflect a prior commitment to the metaphysical hypothesis that there are such things, as it were hovering above and around horse races. More economically, it simply shows us working through the implications of various dispositions to accept and reject betting prices. The 'probability proposition' is a focus for our thoughts about where to put our money. And expressing ourselves in terms of probabilities imposes a necessary logic.[16]

Similarly, if we start with a set of beliefs and attitudes, we can put them into a structured normative space by representing them as beliefs in the ethical proposition. Accepting conditionals and disjunctions shows us working out the implications of various combinations of attitude, or combinations of attitude and belief. We crossed Frege's abyss by creating the ethical proposition, and it is there in order to generate public discourse about which actions to insist upon or forbid, and which attitudes to hold or reject.

A development of this approach has been elegantly presented by Gibbard.[17] Gibbard talks in similar terms of the disjunction: 'Either packing is now the thing to do, or by now it's too late to catch the train anyway', where 'packing is now the thing to do' are words simply expressing the decision to pack. Because decisions can be contested, discussed, disagreed with, we have the indicative form, and we have the devices of propositional logic to keep track of reasoning about decisions. Gibbard develops his own semantics in terms of the notion of a 'hyperdecided state': one in which 'I have a complete view, correct or incorrect, about everything that might be the case in the world, and I would have a universal plan for life, a plan that covers, in detail, every possible situation one could be in that calls for decision'. One can then specify the content of a complex statement involving components expressing decisions by seeing in which hyperdecided states one would not have changed one's mind about them. Logic follows from the conception of consistency: a state of mind is inconsistent if

[16] Why necessary? Because flouting the classical laws corresponds to a disposition to combinations of bets that would lead us to lose whatever happens. This is the 'Dutch book' theorem.

[17] Allan Gibbard, 'A Natural Property Humanly Signified', forthcoming. This presentation is, I think, more perspicuous than that in *Wise Choices, Apt Feelings* because the notion of a completed state is easier to grasp than that of a 'factual-normative' world, which is part of the earlier exposition. But the approach remains essentially the same.

there is no hyperdecided state I could reach from it without changing my mind about something.[18]

Among the most important indirect contexts are those conditionals that express our standards. 'If a person told a lie, then she did something wrong.' Dissent would be dissent from the standard, refusal to be tied to that tree, and we make this refusal if we want to accommodate blameless lying. Among suspect conditionals would be ones expressing what we might think of as a bourgeois morality, which finds rightness or wrongness purely in our gaze: 'If we/enough of us/ . . . think something is wrong, then it is wrong.' This is a tree we would well want to be free from, for we recognize the possibility of correctly dissenting from the herd. Similarly, 'If we had not disapproved of it, it would not have been wrong' only expresses the bourgeois view that our disapprovals actually create the wrongness. Sometimes, this is so. The variable obligations of etiquette, for example, enjoy this status. But usually it is not. Cruelty is not wrong because we disapprove of it, but because it causes pain and anguish.

It is sometimes incorrectly suggested that a 'Humean' must avoid conditionals with factual antecedents and evaluative consequents, because of his embargo on deriving an ought from an is. But this is simply a mistake. Hume can and does vigorously assert the conditionals that express standards. All he denies is that they are truths of reason, having an a priori or analytic status, or one guaranteed by logic alone.

There are other indirect contexts that can appear puzzling. Imagine the case in which you come to me and tell me something using an evaluative term. If you say, for instance, 'If an act creates happiness then it is good,' I will understand you well enough: this is voicing a certain standard, and acknowledging that standard means being disposed to value things on the basis that they create happiness. And values we already have under control. But suppose you are rather more enigmatic, and deploy your values only indirectly. Suppose you say: 'Johnny will do three good deeds by lunch time,' but I don't know what in your estimate counts as a good deed. Can I understand you, when you do not directly express an attitude of evaluation to anything? Since deeds are good in virtue of satisfying standards, and I presume you to understand as much, then I know in principle what will count for you as a vindication of what you say. Johnny will three times perform deeds with a quality X that *you* regard as sufficient to

[18] I would only add here that this requires some preceding notion of consistency in a plan, which would be identified with there existing a possibility of it being realized.



justify an attitude of admiration. I also know what will have to be true for *me* to agree with you. Johnny must perform three deeds with a quality *Y* that *I* regard as sufficient to justify that attitude. *X* may or may not be identical with *Y*. In the upshot, Johnny might perform to your satisfaction but not to mine or vice versa. If the issue is important enough, we are then in disagreement over the standards, and this we express in the propositional way: you think that *X* makes deeds good, and I think *Y* does, and you think Johnny did his three good deeds, while I do not.

For experts in the philosophy of language, it is pleasant to notice that construing such indirect contexts is directly parallel to the construction of indirect contexts in minimalist or 'deflationist' theories of truth.[19] Since these occupy us again in Chapter 9, it is appropriate to introduce them now. According to deflationism, 'is true' is basically a device of disquotation, whose meaning is given by instances of the famous schema:

(T) '*p*' is true if and only if *p*,

where the inverted commas imply mention of a sentence, or in some versions a proposition, or in others an understood-utterance, on the left-hand side, with the same utterance used to talk about whatever it does on the right-hand side. The detailed differences between interpretations of the truth-schema are not important here. But, according to deflationists, (T) encapsulates all we need to know about truth. This is variously expressed by saying either that 'is true' is not a property of utterances, or that it is not a robust or substantive property. It is not one that invites difficult philosophical questions. The idea, in its original formulation by Tarski, is that if we had a sentence of this form for every sentence of a language, and if on the right-hand side we never made use of semantic vocabulary ('is true' or 'refers') then we could, as it were, nevertheless equip a mind that is blind to semantics with enough to give a truth value to every sentence. And this would show that there is no need for such a mind to judge in terms of truth or reference itself in order to be in full cognitive command of the language.[20]

Whether it does show this is controversial. In any event, everyone agrees that something must be added to the schema to enable us to

[19] For an introduction, see my *Spreading the Word* (Oxford: Oxford University Press, 1984), chs. 7 and 8, or Paul Horwich, *Truth* (Oxford: Blackwell, 1990).

[20] Alfred Tarski, 'The Concept of Truth in a Formalized Language', in *Logic, Semantics, Metamathematics* (Oxford: Oxford University Press, 1956).

understand indirect mention, in which the actual proposition asserted remains hidden: 'What Johnny said this morning was true,' for example. We might try a direct quantification: there is a $p$ such that Johnny said '$p$', and $p$.[21] But many philosophers are uncomfortable with this, for it looks as though '$p$' is treated as a variable ranging over propositions at one point, and as an assertion at another. If we avoid this quantification, the neatest way to solve the problem is in terms of an indefinitely large disjunction of conjuncts:

> Either Johnny said '$p_1$' and $p_1$, or Johnny said '$p_2$' and $p_2$, or . . .

where $p_1, p_2, \ldots$ represent the possible things Johnny might have said. It is easy to verify that Johnny will have said something true if, but only if, at least one of these alternatives obtains. So similarly in the present case, when you tell me Johnny will do something good, I can reoresent the content of your saying in terms of a disjunction of conjuncts:

> Either Johnny will do a deed $X_1$ and $X_1$ deeds are good, or Johnny will do a deed $X_2$ and $X_2$ deeds are good, . . .

Interestingly, just as disquotationalists in the theory of truth have some trouble with constructing their list, especially when we turn to consider utterances in a foreign language, we might have some trouble constructing a plausible list of possible standards, especially when we are confronted with the possibility of thoroughly alien standards. So there is an element of idealization involved in supposing that there is some definite list of utterances exhausting the things the reporter might have said, or the list of standards he might have had in mind. But we can well live with that. In each case we would know what is meant by adding 'and so on'. The disjunction sums up that we know what is going on. When someone reports that Johnny did something good, there is some property of deeds that the informant admires, and she believes Johnny did a deed with that property. Notice, however, that the truth of her remark is not simply guaranteed by Johnny doing such a deed: this is the issue of whether $X = Y$ again. I myself expect to be able to voice one of the disjuncts, thereby deploying my standards, in order to regard the remark as true.

With indirect contexts under control, we have now provided all the essential ingredients for a natural account of moral thought. We understand sufficiently what it is to hold something as a value, and to express it using the locutions of belief in a proposition. We understand the ethical proposition as a focus for practical thought.[22]

[21] For experts: this would properly use a substitutional quantifier. The difference does not affect any of the points in the text.



What should a theory of this kind be called? I have called it 'projectivism', but that can sound misleading. It can make it sound as if projecting attitudes involves some kind of mistake, like projecting our emotions onto the weather, or projecting our wishes onto the world by believing things we want to believe. This is emphatically not what is intended. Gibbard calls the view 'expressivism', and I now think that is better. A full-dress title might be 'non-descriptive functionalism', or 'practical functionalism'. In any event, a term I used in my first paper on the subject remains appropriate.[23] There I said that the moral proposition was a 'propositional reflection' of states that are first understood in other terms than that they represent anything, and that remains the core claim. It is the isomorphism between propositional structures and necessary practical states that is the heart of things.

It remains to expand the account, by addressing worries about authority, scepticism, and relativism, and the possibility of moral knowledge, but the core ingredients are now in place.

## 5. REPRESENTATION AND MINIMALISM

There is a sophisticated position in the philosophy of mind and language which would not query the story so far. But it would query whether it makes the ethical proposition anything very special: anything deserving, for instance, special treatment, as the expressivist or projectivist seems to think. The position I have in mind pushes minimalism or deflationism through one more turn. It can usefully be introduced by looking again at the later philosophy of Wittgenstein.

Nobody can deny that Wittgenstein tried to understand many areas of discourse in terms other than those of 'representing the facts'. According to Wittgenstein, whole areas of language that look as if they are dedicated to describing how things are must be understood in other terms. If we think of the areas treated in his later philosophy—the nature of philosophy itself, the nature of self-knowledge, of necessity, of mathematics, of certainty, of religious belief and of ethics—we find that in every case Wittgenstein's approach is to suggest a different

[22] This construction of the ethical proposition is the exercise that I have christened 'quasi-realism'.

[23] 'Moral Realism', reprinted in *Essays in Quasi-Realism* (New York: Oxford University Press, 1994), 111–29.

function for thought in the area. We are doing such things as: laying down rules of grammar, making avowals, making frameworks of description, giving voice to metaphors or the way in which we are seized by a picture, or expressing attitudes. At various times a timid representationalist puts his head over the parapet, and Wittgenstein always dismisses him with a curt reminder of minimalism or deflationism in the theory of truth:

> Someone may say, 'There is still the difference between truth and falsity. Any ethical judgment in whatever system may be true or false.' Remember that '*p* is true' means simply '*p*'. If I say 'Although I believe that so and so is good, I may be wrong': this means no more than that what I assert may be denied.
>
> Or suppose someone says, 'One of the ethical systems must be the right one—or nearer the right one.' Well, suppose I say Christian ethics is the right one. Then I am making a judgment of value. It amounts to adopting Christian ethics. It is not like saying that one of these physical theories must be the right one. The way in which some reality corresponds—or conflicts—with a physical theory has no counterpart here.[24]

The implication is that we can talk of metaphysical, mathematical, modal truth because that is just to repeat our commitments in these areas. But what is going on when we have such a commitment is to be understood in other terms. Wittgenstein often associates this with a verificationist point: we only understand what is meant by making an assertion, or equally by calling it true, in so far as we have a grasp of the procedures that justify or verify the assertion.

Wittgenstein's thought here is the same as that of F. P. Ramsey, whose famous paper 'Facts and Propositions' argues that it is not by staring at the word 'true' that progress is made, but by understanding the function of various kinds of judgement in behavioural terms.[25] Because of the minimalism we can have for free what look like a ladder of philosophical ascent: '*p*', 'it is true that *p*', 'it is really and truly a fact that *p*' . . ., for none of these terms, in Ramsey's view, marks an addition to the original judgement. You can as easily make the last judgement as the first—Ramsey's ladder is lying on the ground, horizontal.[26]

Wittgenstein seemed to think that representation was one thing, but

[24] Rush Rhees, 'Some Developments in Wittgenstein's View of Ethics', *Philosophical Review*, 74 (1965), 23.

[25] F. P. Ramsey, 'Facts and Propositions', *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*, ed. D. H. Mellor (London: Routledge, 1978), 40–57.

[26] See also Chapter 9, where I discuss philosophies that take advantage of the horizontal nature of Ramsey's ladder to climb it, and then announce a better view from the top.



all these different 'language games' did different things. But did he? After all, the minimalism about truth allows us to end up saying 'It is true that kindness is good'. For this means no more than that kindness is good, an attitude we may properly want to express. We can say that the proposition represents the fact that kindness is good. The ethical proposition can be put in the T-schema:

'*X* is good' is true ≡ *X* is good

Anyone understanding the sentence will be prepared to assert the right-hand side if and only if they are prepared to assert the left, in each case voicing the attitude of approval to *X*.

Minimalism seems to let us end up saying, for instance, that 'kindness is good' represents the facts. For 'represents the facts' means no more than: 'is true'. It might seem, then, that our investigations have ended up with a position remarkably like that of Moore. The ethical proposition is what it is and not another thing; its truth means that it represents the ethical facts or the ethical properties of things. We can throw in mention of reality: ethical propositions are really true. Since we already have a sketch of a minimalist theory of ethical cognition, saying that we talk of knowledge that *p* when we are convinced that no improvement has a chance of reversing our commitment to *p*, we might even find ourselves saying that we know moral propositions to be true. Or, really true, or really factually true, or really in accord with the eternal harmonies and verities that govern the universe, if we like that kind of talk. We can add flowers without end: 'it is good to be kind to children' conforms to the eternal normative structure of the world. For this means no more than that it is good to be kind to children.[27] And rather than saying that we hold that it is good to be kind to children, we can if we like say that as we hold that it is good to be kind to children our minds are in harmony with the eternal normative order of things. For this just means the same. I return to this surprising way of looking at things in Chapter 9.

Superficially this might seem like an objection to the investigation, as if the 'quasi-realist' construction has bitten its own tail. It starts from a contrast between expressing belief and expressing an attitude, which it then undermines, by showing how the expression of attitude takes on all the trappings of belief. Since we can handle the ethical proposition exactly like any other, it is not mistaken to say that we voice belief in it, when we do.

[27] 'Eternal' gets in because, I should judge, nothing is going to change to make it other than good to be kind to children.

But in fact this is no objection, and there is no tail-biting. We must remember Wittgenstein's dismissive attitude to invocations of truth and representation when he is dealing with the kinds of commitment that interested him. Just *because* of minimalism about truth and representation, there is no objection to tossing them in for free, at the end. But the commitments must first be understood in other terms. Once more, 'we remain unconscious of the prodigious diversity of all the everyday language games because the clothing of our language makes everything alike'.[28] Our understanding of the kinds of activity involved in specific, ethical, states of mind remain in place, driving the construction of the moral proposition. If we started and finished with a special, *sui generis* representation of moral aspects of the world, we would be drawing a blank. But by starting elsewhere, we can see what is right and justified about finishing saying some of the things Moore did when he spoke in these terms. By getting there this way, however, we have a complete answer to difficulties that destabilize Moore's own package, notably the questions of epistemology, and of why we should be concerned about the ethical properties of things. For we have the answer: what we describe as the ethical properties of things are constructed precisely in order to reflect our concerns.

If we were discontented with minimalism about representation and truth, we might wonder how the story would go with a thick theory of representation to hand. Would that deliver a contrast between 'representing the ethical facts' and 'representing natural facts'? It is hard to say. Obviously there will be some differences between 'ethical facts' and the others. The fact that there is a cannonball on the cushion explains why it is sagging in the middle. The fact that kindness is good explains no such kind of thing. We do not expect laws of ethics to play a role in treatises of physics. Probably the most promising way of finding contrasts would be to think more about the adaptive mechanism that make us sensitive to physical fact, in contrast with the adaptive mechanism that give us an ethical motivational system. The adaptive stories will surely be sufficiently different to give vastly different accounts of 'representation'.

Wittgenstein himself, in the passage quoted above, suggests that with physics 'correspondence' has a different application. But he does not tell us why he thinks this.

We have, however, already met a cluster of considerations that might be drawn from Wittgenstein's later work and that also confuse the picture. We can list:

[28] Wittgenstein, *Philosophical Investigations*, II. xi, p. 224.



(1) The pervasive presence of normativity in any exercise of thought.
(2) The contingent exercise of sensibility involved in apprehending a principle of classification; *hence* perhaps the general impropriety of drawing a fact–value distinction
(3) The *practices* involved in any use of concepts. There are implications for mental action just as much as visible action in the application of any predicate—*hence* the idea that evaluations have a different 'direction of fit' from other representations is blunted.

The drift is that since all exercises of thought, all representation of things as being one way or another, involve evaluation and practice, evaluation should not be thought of as distinct from representation.

But even if this were right, it would not follow that there is anything misdirected about the approach we have adopted. For theorizing about evaluation and values as we have done is quite consistent with supposing that the dispositions of the mind that they consist in are as pervasive as these ideas suggest.

Let us consider an example of the presence of normativity in, say, a simple classification: '*X* is a cat.' Suppose we concede that norms govern the activity of making the judgement. That is, if something is a cat then it is wrong to think of it as any number of things—canine, or insensible, or made of silicon. And it is right to think of it as born of a cat, possessing retractable claws, and so on. To judge that *X* is a cat (sincerely, in full understanding) entails being disposed not to judge that it is a dog, and being disposed to hold that it was born of a cat. Anyone failing to make these moves is apt to forfeit the interpretation. Any judgement has its a priori implications and exclusions. This fact must be couched in normative terms, in terms, that is, of what it is right to conclude, or right to exclude, in light of the proposition.

So now we are back with the idea that in the domain of psychology the normative is everywhere. Psychological states, as explained in the second section of this chapter, relate not by what they typically cause, but what 'makes sense' as an expression of one, or what counts an implication of one, or a reason for one. How exactly does this amount to a strike on behalf of some other theory, and against quasi-realism? Suppose the other theory dislikes e.g. talk of 'attitudes' and likes talk of 'normative beliefs'. Then the two offerings might be:

> (Realism) To believe there is a cat on the mat is to be in a normative space, which is in turn to believe that one ought to . . ., or to be subject to other people believing that one ought to . . ., or to be

disposed in conjunction with other things to believe that one ought to . . .

(Quasi-Realism) To believe there is a cat on the mat is to be in a normative space, which is in turn to have the attitude that one ought to . . ., or to be subject to other people having the attitude that one ought to . . ., or to be disposed in conjunction with other things to have the attitude that one ought to . . .

There is no reason to suppose that the first offering is *ex officio* the default, preferred account. In fact, if anything it looks subject to worrying regresses and circularity (beliefs requiring other beliefs of at least equal and apparently greater complexity). Escape into a different notion, such as sets of norms being embedded in communities and/or psychologies in various ways, alleviates this. Via such a notion we can understand what this normativity amounts to. Without it normativity remains an unmoved mover.

We have to avoid a regress, whereby every belief requires a further background belief about what ought to be held given the first belief. The mistake would be akin to holding that behind every interpretation stands another, rightly opposed by Wittgenstein. To understand the predicate is to *internalize* this system of norms governing its application. It implies a skill, something that can be done better or worse, accurately or ineptly. But it does not imply capacity to articulate a normative structure, capacity to articulate the norms involved in something being a normative structure, and so on.

These reflections show that how we understand the world in one set of terms, or how we understand one part of the world, has normative implications for how it is right to represent it in other terms, or how it is right to understand other parts of it. This is one of the morals of 'holism', although it depends on only a very dilute and uncontroversial version of the doctrine. But it does not break down the distinction between how we understand the world and what we are motivated to do about it. The understanding is practical within the sphere of understanding; our evaluative and ethical life is practical *tout court*.

The upshot is straightforward. In all spheres of thought we make judgements, and judging is subject to standards of correctness. To represent the world as one way or another is to stand ready to be corrected. But in so far as this breaks down a distinction between representation and evaluation, it does so on behalf of evaluation. It is the Apollonian side of the mind, representation and truth, that cannot do without its Dionysian cousin. This only adds importance to the specific investigation of the attitudes, and the other mental dispositions involved in practical reasoning.



It would be possible now to go straight on to explore the structures of motivation and deliberation that we need to understand. Readers impatient to do this should skip the next chapter, which is simply a comparison of the theory I have just presented with other contemporary approaches. This comparison and the criticism of other theories are not essential to understanding the theme of this book. But it may help to locate the exact nature of the view, by better distinguishing it from others with which it could be confused. And it should help to appreciate its strengths if we can maintain a firm grasp of the difficulties in front of alternatives.