

Nietzsche's Minimalist Moral Psychology

Bernard Williams

Nietzsche, Wittgenstein and the extraction of theory

Nietzsche is not a source of philosophical theories. At some level the point is obvious, but it may be less obvious how deep it goes. In this respect, there is a contrast with Wittgenstein. Wittgenstein said repeatedly, and not only in his later work, that he was not to be read as offering philosophical theory, because there could be no such thing as philosophical theory. But his work was less well prepared than Nietzsche's to sustain that position posthumously. There is more than one reason for this.¹ Wittgenstein thought that his work demanded not only the end of philosophical theory but the end of philosophy – something associated, for him, with the end of his demands on himself to do philosophy. That association, of the end of philosophical theory with the end of philosophy, does not deny the idea that if there is to be philosophy, it will take the form of theory; indeed, it readily reinforces that idea. Moreover, the topics on which Wittgenstein wanted there to be no more philosophy – the topics, for him, of philosophy – were traditional topics of academic philosophy. It is not surprising that those who continue theoretical work on those topics still look for elements in Wittgenstein's work itself from which to construct it.

No doubt many who do this lack a suitable irony about what they do to Wittgenstein's texts, but their attitude is not in any important sense a betrayal: less so, in fact, than the attitude of those who think that Wittgenstein did bring to an end philosophical theory on those topics, and themselves sustain an academic activity that consists of reiterating that very thing. Among those who think that there is room for ongoing philosophical theory on those topics, and that Wittgenstein contributed to it, someone owes Wittgenstein an account of why he had ceased to see that this was so. But such an account might be given, and we might come to understand that if Wittgenstein could no longer see the edifice of an intellectual subject, his sightlessness was not that of Samson, but rather that of Oedipus at Colonus, whose disappearance left behind healing waters.

Wittgenstein's posthumous texts, though not designed to express or encourage theory, are not actually mined against its extraction. With Nietzsche, by contrast, the resistance to the continuation of philosophy by ordinary means is built into the text, which is booby-trapped, not only against recovering theory from it, but, in many cases, against any systematic exegesis that assimilates it to theory. His writing achieves this partly by its choice of subject-matter, partly by its manner and the attitudes it expresses. These features stand against a mere

exegesis of Nietzsche, or the incorporation of Nietzsche into the history of philosophy as a source of theories. Some think that these features stand against the incorporation of Nietzsche into philosophy as an academic enterprise altogether, but if that is meant to imply the unimportance of Nietzsche for philosophy, it must be wrong. In insisting on the importance of Nietzsche for philosophy, I mean something that cannot be evaded by a definition of "philosophy". In particular, it cannot be evaded by invoking some contrast between "analytic" and "continental" philosophy. This classification always involved a quite bizarre conflation of the methodological and the topographical, as though one classified cars into front-wheel drive and Japanese, but besides that and other absurdities of the distinction, there is the more immediate point that no such classification can evade the insistent continuities between Nietzsche's work and the business of what anyone calls philosophy. At least in moral philosophy, to ignore them is not simply to adopt a style, but to duck a problem.

I agree with a remark made by Michel Foucault in a late interview, that there is no single Nietzscheism, and that the right question to ask is "what serious use can Nietzsche be put to?" One serious use is to help us with issues that press on any serious philosophy (in particular, moral philosophy) that does not beg the most basic of its own questions. Nietzsche will not help if he is taken to impose some one method on us. I have already said that I find his texts securely defended against exegesis by the extraction of theory; but it does not follow, and it is important that it does not follow, that when we are trying to put him to serious use our philosophy should contain no theory. This is because the insistent continuities between his questions and our business run in both directions. Some of the concerns to which he speaks are going to be better met – that is to say, met in a way in which we can better make something of them – by quite other styles of thought, and perhaps by some theory of other origins: certainly not by theoretical, or again anti-theoretical, incantations supposedly recovered from Nietzsche himself.

Naturalism and realism in moral psychology

There is some measure of agreement that we need a "naturalistic" moral psychology, where this means something to the effect that our view of moral capacities should be consistent with, even perhaps in the spirit of, our understanding of human beings as part of nature. A demand expressed in some such terms is perhaps accepted by most philosophers, apart from some *anciens combattants* of the wars of freewill. The trouble with this happy and extensive consensus, however, and no doubt the condition of it, is that no-one knows what it involves. Formulations of the position tend to rule out too much or too little. The position rules out too much if it tries reductively to ignore culture and convention; this is misguided even on a scientific basis, in the sense that to live

under culture is a basic part of the ethology of this species.² It rules out too little if it includes many things that have been part of the self-image of morality, such as certain conceptions of moral cognition; a theory will scarcely further the cause of naturalism in this sense if it accepts as a basic feature of human nature the capacity to intuit the structure of moral reality. It is tempting to say that a naturalistic moral psychology explains moral capacities in terms of psychological structures that are not distinctively moral. But so much turns on what counts as explanation here, and what it is for a psychological element to be distinctively moral, that it remains persistently unclear whether the formula should be taken to be blandly accommodating, or fiercely reductive, or something in between.

The difficulty is systematic. If a "naturalistic" moral psychology has to characterise moral activity in a vocabulary that can be applied equally to every other part of nature, then it is committed to a physicalistic reductionism that is clearly hopeless. If it is to describe moral activity in terms that can be applied to something else, but not everything else, we have not much idea what those terms may be, or how "special" moral activity is allowed to be, consonantly with naturalism. If we are allowed to describe moral activity in whatever terms moral activity may seem to invite, naturalism excludes nothing, and we are back at the beginning. The trouble is that the very term "naturalism" invokes a top-down approach, under which we are supposed already to know what terms are needed to describe any "natural" phenomenon, and we are invited to apply those terms to moral activity. But we do not know what those terms may be, unless they are (uselessly) the terms of physics, and this leads to the difficulty.

In this quandary, we can find in Nietzsche both a general attitude, and some particular suggestions, that can be a great help.³ I shall say something later about what I take some of his suggestions to be. The general attitude has two relevant aspects, which have to be taken together. First, to the question "how much should our accounts of distinctively moral activity add to our accounts of other human activity?" it replies "as little as possible", and the more that some moral understanding of human beings seems to call on materials that specially serve the purposes of morality – certain conceptions of the will, for instance – the more reason we have to ask whether there may not be a more illuminating account that rests only on conceptions that we use anyway elsewhere. This demand for moral psychological minimalism is not, however, just an application of an Occamist desire for economy, and this is the second aspect of the Nietzschean general attitude. Without some guiding sense of what materials we should use in giving our economical explanations, such an attitude will simply fall back into the difficulties we have already met. Nietzsche's approach is to identify an excess of moral content in psychology by appealing first to what an experienced, honest, subtle and unoptimistic interpreter might make of human behaviour elsewhere. Such an interpreter might be said to be – using an obviously and unashamedly evaluative expression – "realistic", and we might say that what this approach leads us towards is a realistic, rather than a naturalistic, moral psychology. What is at issue is not the application of an

already defined scientific programme, but rather an informed interpretation of some human experiences and activities in relation to others.

Such an approach can indeed be said to involve, in Paul Ricoeur's well-known phrase, a "hermeneutics of suspicion". As such, it cannot compel demonstratively, and does not attempt to do so. It invites one into a perspective, and to some extent a tradition (one marked by such figures as Thucydides, for instance, or Stendhal, or the British psychologists of morals whom Nietzsche described as old frogs), in which what seems to demand more moral material makes sense in terms of what demands less. The enterprise can work, however, only to the extent that the suspicion it summons up is not a suspicion of everything. Writers on Nietzsche typically pay most attention to his claims, or what appear to be his claims, that every belief about the relations of human beings to reality are open to suspicion, that everything is, for instance, an interpretation. Whatever may need to be said at that level, it is equally important that when he says that there are no moral phenomena, only moral interpretations, a *special* point is being made about morality. This does not mean that we should simply forget, even in these connections, the larger claims. We need to get a deeper understanding of where these points of particular suspicion are to be found, and it may be helpful to work through larger claims on a path to getting a grasp on more limited claims. This is particularly so if we bear in mind that "claim", for Nietzsche, is in fact rarely the right word. It is not only too weak for some things he says and too strong for others; we can usefully remember, too, (or perhaps pretend) that even when he sounds insistently or shrilly expository, he is not necessarily telling us something, but urging us to ask something.

In the rest of this paper, I shall try to assemble some of Nietzsche's suggestions about a supposed psychological phenomenon, that of willing. I shall leave aside many interesting things that Nietzsche says about this concept, in particular about its history. My aim is to illustrate, through a schematic treatment of this central example, the way in which a method of suspicion – the search, one might almost say, for a culprit – can help us to achieve a reduced and more realistic moral psychology.

Illusions of the self

Speaking seriously, there are good reasons why all philosophical dogmatizing, however solemn and definitive its airs used to be, may nevertheless have been no more than a noble childishness and tyronism. And perhaps the time is at hand when it will be comprehended again and again *how little* used to be sufficient to furnish the cornerstone for such sublime and unconditional philosophers' edifices as the dogmatists have built so far; any old popular superstition from time immemorial (like the soul superstition which, in the form of the subject and ego superstition, has not even yet ceased to do mischief); some play

on words perhaps, a seduction by grammar, or an audacious generalization of very narrow, very personal, very human, all too human facts.⁴

The general point that Nietzsche makes here (one shared with Wittgenstein, and indeed J. L. Austin, about the extraordinary lightness of philosophical theories) is directed to a particular idea, that the ego or self is some kind of fiction. Later in the same book he follows Lichtenberg in criticising the *cogito* as the product of grammatical habit. Elsewhere, he makes a similar point more specifically about action. He quotes a sceptic:

“I do not in the least know what I am doing. I do not in the least know what I ought to do.” You are right, but be sure of this: *you are being done* [*du wirst getan*], in every moment. Mankind at all times has mistaken the passive for the active: it is their constant grammatical mistake.⁵

Many ideas might be drawn from this complex, some of them uninviting; for instance, that we never really do anything, that no events are actions. More interestingly, Nietzsche can be read as saying that *action* is a serviceable category of interpretation, but a local or dispensible one. This seems to me hardly less implausible, but some have accepted it.⁶ If people perform actions, then they perform them because they think or perceive certain things, and this is enough to dispose, further, of a crude epiphenomenalism which might be found in some of Nietzsche’s sayings – perhaps in his suggestion that all action is like willing the sun to rise when the sun is just about to rise.

Nietzsche’s doubts about action are more usefully understood, I suggest, as doubts not about the very idea of anyone’s doing anything, but rather about a morally significant interpretation of action, in terms of the will. The belief in the will involves for him two ideas in particular: that the will seems something simple when it is not, and that what seems simple also seems to be a peculiar, imperative, kind of cause.

Philosophers are accustomed to speak of the will as if it were the best-known thing in the world . . . But . . . [w]illing seems to me to be above all something *complicated*, something that is a unit only as a word – and it is precisely in this one word that the popular prejudice lurks, which has defeated the always inadequate caution of philosophers.⁷

He goes on to explain that what is called “willing” is a complex of sensations, thinking, and an affect of command. He points to the consequences of our being both the commanding and the obeying party, and of our “disregarding this duality”.

Since in the great majority of cases there has been an exercise of will only when the effect of the command – that is, obedience; that is, the action – was to be *expected*, the appearance has translated itself into the feeling,

as if there were a *necessity of effect*. In short, he who wills believes with a fair amount of certainty that will and action are somehow one; he ascribes the success, the carrying out of the willing, to the will itself, and thereby enjoys an increase of the sensation of power that accompanies all success.

What exactly is the illusion that Nietzsche claims to expose here? It is not the idea that a certain experience is a sufficient cause of an action. He does indeed think that the experiences involved in "willing" do not reveal, and may conceal, the shifting complex of psychological and physiological forces that lies behind any action, the constant, unknown, craving movements that make us, as he puts it, a kind of polyp.⁸ But it is not that the experience sets itself up as the cause. Rather, the experience seems to reveal a different kind of cause, and suggests that the cause does not lie in any event or state of affairs – whether an experience of mine or otherwise – but in something that I refer to as "I". Such a cause seems to be related to the outcome only in the mode of prescription, through an imperative; and since this stands in no relation to any causal set of events, it can seem to bring about its outcome *ex nihilo*.

Of course, any sensible theory of action, which allows that there is indeed action, and that thoughts are not merely epiphenomenal in relation to it, will have to allow that the consciousness of acting is not the same as a consciousness that a state of mine causes a certain outcome. This follows merely from the point that the first-personal consciousness which one has when involved in action cannot at the same time be a third-personal consciousness of that very involvement. But the first-personal consciousness which an agent necessarily has does not in itself have to lead to the kind of picture that Nietzsche attacks; action does not necessarily involve this understanding of itself.⁹ The picture is a special one, particularly associated with a notion such as "willing", and when it is present, it is not merely a philosophical theory of action, but can accompany many of our thoughts and moral reactions. So where does it come from, and what does it do?

Part of Nietzsche's own explanation is to be found in the course of one of his most famous passages:

For just as the popular mind separates the lightning from its flash, and takes the latter for an *action*, for the operation of a subject called lightning, so popular morality also separates strength from expressions of strength, as if there were a neutral substratum behind the strong man, which was *free* to express strength or not to do so. But there is no such substratum; there is no "being" behind doing, effecting, becoming; "the doer" is merely a fiction added to the deed – the deed is everything. The popular mind in fact doubles the deed; when it sees the lightning flash, it is the deed of a deed: it posits the same event first as cause and then a second time as its effect.¹⁰

There are two helpful ideas in this account. One is that the picture under attack involves a kind of double counting. The self or I that is the cause is ingenuously introduced as the cause of an action. If my agent-self produces only a set of events, it may seem that I shall not have enough for my involvement in the action: I shall be at best the "pilot in a ship" to which Descartes referred. The same result follows from the idea that the mode of causation is that of command: obedience to a command consists of an action. But commanding is itself an action. The self can act (at one time rather than another, now rather than earlier) only by doing something – the thing it does, willing. In making action into something that introduces an agent-cause, the account has a powerful tendency to produce two actions.

The second helpful thought to be recovered from Nietzsche is that such a peculiar account must have a purpose, and that the purpose is a moral one.

The target of blame

The purpose of the account can be read from the way in which it associates two ideas, which contribute to its incoherence and together compound it. One idea is that there is a metaphysically special unit, a real action, unlike anything else that can be individuated among the world's processes. The other is that this stands in an unmediated relationship – something like being an effect *ex nihilo* – to something of quite a different kind, again unique – a person, or self, or agent. There is an idea that needs items standing in just such a relation: it is a certain purified conception of blame.

Blame needs an occasion – an action – and a target – the person who did the action and who goes on from the action to meet the blame. That is its nature; as one might say, its conceptual form. It does not need these things in the pure and isolated form implied by the account of the will. The Homeric Greeks blamed people for doing things, and whatever exactly went into their doing so, it was not all this. Rather, this version of the occasion and the target will be demanded by a very purified conception of blame, a conception seemingly demanded by moral justice. It is important that the mere idea of just compensation does not make this demand, nor every idea of responsibility. If A has been damaged by B's careless action, B may be held responsible for the loss and reasonably required to compensate A, though the loss to A formed no part of what B willed. A very exact concentration on B's will, and the purely focussed conception of blame that goes with it, are demanded not merely by responsibility or demands in justice for compensation, but by something more specific.

It is not hard to find an explanation of the more specific demand. It lies in the seeming requirement of justice that the agent should be blamed for no more and no less than what was in his power. What the agent brought about (and for which, in the usual order of things, he may be asked to provide compensation) may very well be a matter of luck, but what he may be strictly (as these conceptions say, "morally") blamed for cannot be a matter of luck, and must

depend in a strict and isolable sense on his will. It is appropriately said that what depends on his will is what is strictly *in his power*: it is with regard to what he wills that the agent himself has the sense of power in action to which Nietzsche refers. As agents, and also as blamers under justice, we have an interest in this picture.

The needs, demands and invitations of the morality system are enough to explain the peculiar psychology of the will. But there is more that needs to be said about the basis of that system itself. Nietzsche himself famously suggested that a specific source for it was to be found in the sentiment of resentment – a sentiment which itself had a historical origin, though hardly one that he locates very precisely. I shall not pick up the historical aspect, but I think it is worth suggesting a brief speculation about the phenomenology of focussed blame, which is a close enough relation to Nietzsche's "genealogy", perhaps, to be a version of it.¹¹

If there is a victim with a complaint for a loss, there is an agent who is to blame, and an act of that agent which brought about the loss. The anger of the victim travels from the loss to the act to the agent; and compensation or recompense given by the agent will acknowledge both the loss and the fact that he was the cause of the loss. Suppose the agent brings about a harm to the victim, and does so intentionally and voluntarily; where "intentionally and voluntarily" is not supposed to invoke the special mechanisms of the will, but merely means that the agent knew what he was doing, wanted to do it, and was in a normal state of mind when he did it. Suppose that the agent is not disposed to give compensation or reparation, and that the victim has no power to extract any such thing from him. In refusing reparation, the agent refuses to acknowledge the victim or his loss; it is a peculiarly vivid demonstration of the victim's powerlessness.

These circumstances can give rise, in the victim or in someone else on behalf of the victim, to a very special fantasy of retrospective prevention. As victim, I have a fantasy of inserting into the agent an acknowledgement of me, to take the place of exactly the act that harmed me. I want to think that he might have acknowledged me, that he might have been prevented from harming me. But the idea cannot be that I might in some empirical way have prevented him: that idea presents only a regret that it was not actually so and, in these circumstances, a reminder of humiliation. The idea has to be, rather, that I, now, might change the agent from one who did not acknowledge me to one who did. This fantasied, magical, change does not involve actually changing anything, and it therefore has nothing to do with what, if anything, might actually have changed things. It requires simply the idea of the agent at the moment of the action, of the action that harmed me, and of the refusal of that action, all isolated from the network of circumstances in which his action was actually embedded. It involves precisely the picture of the will that has already been uncovered.

Much can grow from this basic feeling. It lays the foundation for the purest and simplest construction of punishment, and it is very significant how the language of retribution naturally deploys teleological notions of conversion,

education, or improvement (“teaching him a lesson”, “showing him”) while insisting at the same time that its gaze is entirely retrospective and that, inasmuch as it is purely retributive, it does not look to actual reform.¹² But the construction is at least as much at work when it is not a question of any actual punishment, but only of the purely moral conceptions of guilt and blame, and it then involves a further abstraction; it introduces not only retribution’s idea of retrospective causation, but morality’s idea of an authoritative but sanctionless law, of a judgement that carries no power except that judgement itself.

Conclusion

This is, of course, only a sketch of a possible account, drawn (fairly directly) from Nietzschean materials. The most important feature of it, for the present purpose, is its structure. We start with a supposed psychological phenomenon, willing, associated with the conception of the self in action. The phenomenon seems recognizable in experience, and it seems also to have a certain authority. Its description already presents difficulties and obscurities, but proposals merely to explain it away or to ignore it seem typically to have ignored something important, even to leave out the essence of action. Reminded both that different pictures of action have been held in other cultures, and that the notion of action itself is less than transparent, we can be helped to see that the integrity of action, the agent’s genuine presence in it, can be preserved without this picture of the will. The process by which we can come to see this may be complex and painful enough for us to feel, not just that we have learned a truth, but that we have been relieved of a burden.

Since the picture is neither coherent nor universal, yet has this authority, we need to ask where it comes from and what it does. It is not itself manifestly tied to morality, offering rather a picture of voluntary action in general, but there is a moral phenomenon, a certain conception of blame, which it directly fits. This conception, again, is not universal, but is rather part of a special complex of ethical ideas which has other, and related, peculiar features. The fit between the special psychological conception and the demands of morality enables us to see that this piece of psychology is itself a moral conception, and one that shares notably doubtful features of that particular morality itself. In addition to this, we may be able to supply some further psychological conceptions which help us to understand the motivations of this particular form of the ethical. Those conceptions, as presented by Nietzsche under the name of resentment, certainly lead out of the ethical altogether, into the categories of anger and power, and it cannot be a matter simply for philosophy to decide how much those categories will explain. Other explanations may be needed, and it may be that they will prove to be more basically linked to notions of fairness, for instance. But in laying such explanations against one another, and in diagnosing the psychology of willing as a demand of the morality system itself, we shall be

following a distinctively Nietzschean route towards the naturalization of moral psychology.

Bernard Williams
Corpus Christi College, Oxford

NOTES

¹ Even when we leave aside the point that there is only one work by Nietzsche (*The Will to Power*) that is not a work by Nietzsche, whereas the later works of Wittgenstein are, as whole books, very variably his.

² I have discussed this point at greater length in "Making Sense of Humanity".

³ It will be obvious that Nietzsche's interest is located by the present discussion mostly in terms of his more "sceptical" works, rather than in (for instance) his ideas of self-overcoming. This is not to deny that they, too, can have their uses. In any case, there is no hope of getting anything from those ideas without setting them against his accounts of familiar morality.

⁴ *Beyond Good and Evil*, Preface. The reference to Lichtenberg, below, is at sec. 17.

⁵ *Daybreak*, 120. The passage about the sunrise, mentioned below, is also from *Daybreak*, 124.

⁶ E.g. Frithjof Bergmann, "Nietzsche's Critique of Morality". Bergmann includes "individual agency" (along with such items as selfhood, freedom and guilt) in a list of concepts allegedly special to our morality; he takes himself (wrongly, I think) to be following Clifford Geertz in the claim that it was not known in traditional Bali. Similar errors have been made about the outlook of Homeric Greece: see below, note 9.— The idea that *action*, in our ordinary understanding of it, is a dispensible and indeed mistaken conception is shared by a very different kind of philosophy, eliminative materialism; in that case for scientific reasons.

⁷ *Beyond Good and Evil*, 19. The whole section is relevant.

⁸ *Daybreak*, 119.

⁹ This is clearly illustrated by the treatment accorded by some scholars to the Homeric conception of action; not finding in Homer this picture of action, they have thought that the archaic Greeks either had no idea of action, or had an imperfect one, lacking the concept of the will. I discuss this and related misconceptions in *Shame and Necessity*: see in particular chapter 2.

¹⁰ *The Genealogy of Morals*, First Essay, sec. 13.

¹¹ A Nietzschean genealogy typically combines, in a way that analytical philosophy finds embarrassing, history, phenomenology, "realistic" psychology, and conceptual interpretation. The historical stories, moreover, strikingly vary from one context to another. Some of Nietzsche's procedures are to be seen specifically in the light of Hegel's Phenomenology, and of his recurrent amazement that there could have been such a thing as Christianity. Some are certainly less helpful than others. But the mere idea that we need such elements to work together is surely right. We need to understand what parts of our conceptual scheme are, in what degree, culturally local. We understand this best when we understand an actual human scheme that differs from ours in certain respects. One, very important, way of locating such a scheme is finding it in history, in particular in the history of our own. In order to understand that other scheme, and to understand why there should be this difference between them and us, we need to understand it as a

human scheme; this is to understand the differences in terms of similarities, which calls on psychological interpretation. Very roughly speaking indeed: a Nietzschean genealogy can be seen now as starting from Davidson plus history.

¹² A particularly illuminating example is Robert Nozick's discussion of retributive punishment in *Philosophical Explanations*, p. 363 seq. His heroic attempt to express what pure retribution tries to *achieve* (as opposed to what, in actual fact, it does) reveals, it seems to me, that there is no logical space for it to succeed.

REFERENCES

- Bergmann, F. (1988), "Nietzsche's Critique of Morality", in Robert C. Solomon and Kathleen M. Higgins (eds.), *Reading Nietzsche*. Oxford: Oxford University Press.
- Nietzsche, F., *Beyond Good and Evil*. Translated by Walter Kaufmann. New York, 1966.
- Nietzsche, F., *Daybreak*. Translated by R. J. Hollingdale. Cambridge, 1982.
- Nietzsche, F., *The Genealogy of Morals*. Translated by Walter Kaufmann and R. J. Hollingdale. New York, 1967.
- Nozick, R. (1984), *Philosophical Explanations*. Oxford: Oxford University Press.
- Williams, B. (1991)., "Making Sense of Humanity", in James J. Sheehan and Morton Sosna (eds.), *The Boundaries of Humanity: Humans, Animals, Machines*. California University Press.
- Williams, B. (1993), *Shame and Necessity* (the Sather Classical Lectures, 1989). California University Press.