Introduction

Despite being somewhat long in the tooth at the time, Aristotle, Hume and Kant were still dominating twentieth century moral philosophy. Much of the progress made in that century came from a detailed working through of each of their approaches by the expanding and increasingly professionalized corps of academic philosophers. And this progress can be measured not just by the quality and sophistication of moral philosophy at the end of that century, but also by the narrowing of some of the gaps between Aristotelian, Humean and Kantian philosophers.

The most significant legacies from the nineteenth century were utilitarianism represented by Bentham and Mill, and by contrast the nihilism of Nietzsche. The hold that utilitarianism had over the discourse remained practically undiminished despite the numerous attempted refutations. And coming from a completely different philosophical perspective Nietzsche had generated an intoxicating mixture of scepticism, mysticism and the promise of freedom from conventional morality.

All this was worked through the great twentieth century philosophical movements like positivism, linguistic philosophy and existentialism. And in subtler ways the great cultural shifts of the twentieth century left their mark too. What became described as agent-centred morality reflected the reaction to the various threats of industrial mechanisation, self-righteous genocide, totalitarian politics and paternalism. And in addition to the abiding interest in the role of ethics in politics and law, moral philosophy developed a concern with the role of ethics in medicine, business and our relationship with the environment.

The big questions that concerned twentieth century moral philosophers were largely inherited from previous ages. Very roughly they fell into five categories: (i) the conceptual – how we should understand ethical language; (ii) the metaphysical - whether ethical value is real; (iii) the epistemological - how moral judgements are
grounded or justified; (iv) the moral psychological – how morality is related to the will; and (v) the normative – what, if any, are the proper structuring principles of an ethical system. I will deal with these roughly in turn, but not try to squeeze every aspect of twentieth century moral philosophy into a particular category. Most work in ethics has addressed several of these questions simultaneously.

That there was any answer to the conceptual question of what ‘good’ means was challenged early on by G.E. Moore’s attack on the naturalistic fallacy. This was supposed to be the fallacy of providing a definition of the evaluative word, ‘good’, in non-evaluative terms. Moore’s attack did not stop a series of ‘naturalistic’ attempts to answer the question (though not always in terms of a definition), from John Dewey and R.B. Perry in the early part of the century through to the Cornell realists towards the end of it.

More significant perhaps was the discovery by so-called ‘non-cognitivist’ philosophers like Charles Stevenson and Richard Hare that an account of the expressive significance of moral language might be attempted while denying that moral terms had any descriptive meaning as such. The non-cognitivists were following up David Hume’s famous dictum that morality was not derived from reason alone, arguing instead that it was merely an expression of individual or social attitudes. Expressing these attitudes was seen to be distinct from describing them. Just as for example frowning may be an expression of disapproval but not a description of disapproval, so it was claimed moral utterances were merely expressions and not descriptions.

The metaphysical questions about the reality of moral values and the factuality of moral claims rumbled on throughout the century. Non-cognitivists generally argued that there were no moral facts corresponding to ethical judgements. But then so did some cognitivists like Gilbert Harman, who thought that moral judgements were derived from reason but just never objectively true. With advances in the philosophy of language the issue became subtler. Moral realists did not after all have to be committed to some Platonic realm in which moral values dwelled or to the existence of natural or unnatural properties in virtue of which moral claims had objective truth. They just needed to argue that the norms for the correctness of moral judgements were in some cases not dependent on the attitudes of the individual or of their cultural community.
Intuitionism was one way to answer the epistemological question for realist philosophers who had rejected a naturalistic answer to the conceptual question. They claimed that we know the answers to moral questions directly and not through a process of inference. This view, which may have started off looking horribly like hand-waving, has turned out to have a real force as it developed through the work of G.E. Moore, H.A. Prichard, W.D. Ross, and later philosophers like Jonathan Dancy and Robert Audi.

This kind of intuitionism was often associated with an Aristotelian approach to moral psychology in which moral judgements, rationality, emotions and action are related to one another. Oxford philosophers like Iris Murdoch and John McDowell were particularly influential in developing this sensibility approach to ethics, though the role of emotions in ethical thinking had been examined much earlier in the century by continental philosophers like Max Scheler.

The alternative, developed by Kantian philosophers like Stephen Toulmin and Alan Gewirth was to say that we know the answers to moral questions through a process of reasoning. The central issue in moral psychology for such Kantians is the relationship between moral judgement, rationality and action, with emotional attitudes having less significance.

With the fifth of these big questions – normative ethics - there was no agreement as to whether the abstract discipline of philosophy could really say anything useful. The rather tired distinction between consequentialist and deontological approaches remained stubbornly in play throughout the century. Consequentialism was the view that the rightness or wrongness of an action should be determined by the goodness or badness of its consequences or expected consequences. Utilitarianism was the classic example of a consequentialist approach. Deontology, represented paradigmatically by Kantian ethics, was the view that the rightness or wrongness of an action should be determined by whether or not the action conformed to proper principles of morality – principles which made explicit the duties of the agent.

In some taxonomies virtue theory was added to this pair. According to virtue theory, represented by philosophers like Philippa Foot (1978) and Michael Slote (1992), it was
important to consider the qualities of the agent in assessing the rightness or wrongness of their actions. This Aristotelian approach to ethics was usually associated with the so-called ‘particularist’ view that there were no true universal principles that could determine the rightness or wrongness of an action. Having a properly virtuous character, and thus having a properly attuned moral sensibility, meant that one was sensitive to the moral considerations that applied in a situation in its infinite particularity. Any general principle would be bound to ignore important moral distinctions. In continental Europe this particularist approach was developed by philosophers in the phenomenological tradition like Martin Buber and Emanuel Levinas.

**Moore’s Challenge**

The century began more or less with G.E. Moore’s (1903) broadside on Mill and others whom Moore accused of trying to derive an understanding of what was good from a basic definition of ‘good’ in morally neutral terms. In other words Moore was rejecting a reductive analysis of ‘good’.

Moore distinguished definitions of ‘good’ from extensional characterisations of ‘good’. He took it to be a useful task of philosophical ethics to spell out the extension of the predicate – i.e. to determine the set of all things that are good. And it was also a useful task to characterize this extension in various illuminating ways – to provide a sort of taxonomy of goodness. But he took it to be an error to try to define the predicate – i.e. to try to capture its meaning - in terms of other predicates. In particular it would be a mistake – and this is what he called the naturalistic fallacy - to do an extensional analysis and then think that one had captured the meaning of the predicate. This would be like observing that all yellow things reflected light of a certain frequency range and then to conclude that the meaning of the word ‘yellow’ was ‘reflects light in such and such a frequency range’.

If the meaning of the term ‘yellow’ were given by this sort of characterization of its extension then the claim that yellow things reflect light in such and such a frequency range would be empty and pointless, since it would just mean that things that reflect light in such and such a frequency range reflect light in such and such a frequency range.
range. And the same goes for an attempt to define ‘good’ as ‘desired’ or as ‘involving pleasure’ for instance. To say that things which are desired are good would be to say nothing significant if it were just a statement of a definition of the term ‘good’.

Moore proposed a test – the open question argument - for whether a statement might count as providing the meaning of a predicate. If some statement just tells us what we mean by the term and that statement is fully understood then it is not open to further questioning. Suppose I ask whether vixens are foxes, and I am told correctly that it is part of what we mean by the word ‘vixen’ that it is a female fox. This closes off the question; no further discussion is needed. But if I ask whether good things are things that give people pleasure, the question remains open however much I know the meaning of the word ‘good’. So it cannot be part of the meaning of the word ‘good’ that good things give people pleasure.

This test has to be applied with great care however since meaning itself is not always completely transparent. It may seem to be an open question just what some word means even when one can use the word reasonably well in most contexts.

An extension of Moore’s open question argument that was applied by various philosophers through the course of the century depended on accepting David Hume’s internalism – his claim that motivation is internal to moral judgement. Given this, you cannot truly judge something to be right or good without thereby having some motivation to act or feel accordingly. If that is right there should be no room for the following question: “I can see that this is good, but why should I care?” And likewise, for any attempted analysis, A, of ‘good’, there should be no room for the question: “I can see that this is A; but why should I care?” For example, in response to an analysis of good in terms of conventional acceptance I might say: “I can see that obeying the law is conventionally accepted, but why should I care about that.” To the extent that the question makes sense the putative analysis fails.

The diagnosis of the naturalistic fallacy has been ridiculed in approximate proportion to how incredibly influential it has been.¹ It is common to say that the naturalistic fallacy

¹ An important early attack came from William Frankena (1939).
is neither naturalistic nor a fallacy. To say it is not a fallacy is to say there is no mistake made in confusing definitions and extensional characterisations. And saying this might be justified by accepting Quine’s (1951) attempted demolition of the analytic/synthetic distinction. But I think that would be an anachronism. Moore was criticising attempts to provide analyses of our moral concepts (although the extent to which such criticism hit the mark with Mill is questionable). And Moore’s claim that though such accounts have said useful things about the extension of the predicate ‘good’, they failed to provide analyses of meaning, fits happily with Quine’s rejection of the notion of analytic truth. Quine was rejecting the idea of providing analyses generally; Moore was rejecting the idea of providing an analysis of ‘good’.

**Naturalistic Ethics**

Certainly, however, it may seem odd to call the fallacy naturalistic, especially since Moore himself recognised that the fallacy could also apply to non-naturalistic definitions. Roughly speaking, a natural quality was taken to be a quality that is either present in ordinary experience or is the sort of quality that figures in natural science. Moore thought that goodness was neither of these, and consequently that no definitional analysis of ‘good’ could be had in purely naturalistic terms. Nevertheless someone with a perfectly good naturalistic extensional characterisation of goodness might fallaciously take themselves to have a naturalistic definition of goodness. Taking yourself to have a naturalistic definition of goodness is not itself a fallacy; it is just a mistake. The fallacy is what might lead you to this mistake.

With the sustained attack on the fact/value distinction, made especially by the sensibility theorists of the second half of the century, the natural/non-natural distinction would seem less clear. In particular it was no longer so clear that goodness could not figure in ordinary experience. But Moore’s argument retains its force, since it was aimed specifically at those who sought a reductive definition of the moral notion of goodness in non-moral terms. His opponents were those who worked with the distinction between the non-natural and the natural and then tried to define one in terms of the other.

The sorts of things that might have been taken to count as naturalistic sources for our moral concepts are individual preferences, individual feelings of desire, enjoyment,
pleasure, etc, individual emotions, individual judgments of approbation, individual acts or dispositions of will, social agreements and rules, and cultures of praise and blame. But many naturalistic approaches to ethics were not subject to Moore’s attack since they did not attempt to define moral terms in such naturalistic terms. Generally, in the spirit of Aristotle, they sought to ground our use of moral language in naturalistic terms, not by reducing the moral to the natural but by anchoring the moral to the natural. So Aristotle himself, who took a virtuous character to be one that fulfilled human nature, certainly did not take the idea of virtue to be reducible to that of human nature.

And one of the twentieth century’s most influential naturalists, John Dewey, argued that value is connected with experiences of enjoyment, but only in the sense that these experiences provide the material that is elaborated in the process of rational discourse to construct value. He argued that morality should not float free of human experience as it did in what he took to be the empty rationalism of Kant. But he was just as concerned to avoid the sensationalist approach which identified morality with human experience and so left it useless as a means for providing direction to conduct.

Very much in the spirit of both Aristotle and Hume he claimed that the formation of taste is the chief matter wherever values enter in. “The formation of a cultivated and effectively operative good judgement or taste with respect to what is esthetically admired, intellectually acceptable and morally approvable is the supreme task set to humans by the incidents of experience.” (1929, 209)

Dewey took the experimental method that is acknowledged as the way to arrive at scientific truth to be the way to arrive at moral truth. Moral rules and principles were hypotheses to be tested in the imagination and in practice. Only by trying out particular moral stances and adapting and refining them in response to their success or failure at guiding in a clear and settled way could these stances be validated. This was an early example of what became known in moral philosophy following Rawls (1951 and 1971) as the method of reflective equilibrium.

---

2 Some of Dewey’s most significant work on ethics was in the 1920s with his *Human Nature and Conduct* (1922) and “The Construction of Good”, which was chapter 10 of his *The Quest for Certainty* (1929)
Indeed the practice of interacting with situations in order to test moral hypotheses for stability and clarity not only determines the correctness of these hypotheses, it also constructs the correctness of these hypotheses, according to Dewey. A situation does not come to us with a solution that we are initially too puzzled to find. By interacting with the situation we make it morally determinate. We invest it with rightness and wrongness.

This might be taken in a weaker or stronger way. The weaker way would be to observe that a moral situation is not just a blankly external state of affairs. It is a set of questions and problems, which depend essentially on how the state of affairs is conceptualised and interacted with. By engaging imaginatively with a moral situation we can describe it in terms of alternative possible courses of action. There is no canonical description of a state of affairs that fixes the moral issues and dilemmas. So in this sense we might be said to construct the goodness in a situation.

The stronger sense would take the moral truth of a situation to be a function of what we do with it in our process of rational interaction with it. Something has value if it is associated with the feeling of enjoyment at the end of the rational process of experimentation guided by the goal of “cultivated and effectively operative good judgement and taste”. When you have interacted with a situation in the designated way and as a result object X is judged to have value, then that is what it is for object X to have value.

This stronger sense of the idea of the construction of good might be vulnerable to Moore’s naturalistic fallacy argument. Moore would pose the open question: is the thing that emerges with a judgement of value from such a process really good? And if this is genuinely an open question, then the notion of goodness is not captured by this definition. But it is by no means clear that this idea that the goodness of something is a function of your feelings or attitudes is the right claim to pin on Dewey. He was usually very cautious about making such a strong claim. And if he was not making any attempt to define ‘good’ in his notion of the construction of good, then he was not vulnerable to Moore’s naturalistic fallacy argument.
Perhaps the very fact that there are so few twentieth century approaches to moral philosophy that can be properly criticised as committing the naturalistic fallacy should be taken to be a testament to the influence of Moore’s argument rather than, as it is often taken, an indication of its irrelevance. The naturalistic definition of colour that Moore took to reveal more clearly the absurdity of naturalistic definitions of non-natural terms has turned out to have many more adherents than naturalistic definitions of goodness. And this is may be because philosophers of colour did not feel they had Moore breathing down their necks.

One approach that was often taken to be subject to Moore’s attack at the time was that of E. Westermarck (1906-08). This was a sociological, psychological and evolutionary approach; and since Westermarck was not really a philosopher he was an easy target. Westermarck defined the wrongness of an action in terms of an individual’s inclination to disapprove of it and sometimes too of a tendency across society to disapprove of it. Moore’s argument presents this account with the question of whether it is wrong to do something that you or society are inclined to disapprove of. If this question is genuinely an open question – one whose answer does not just fall out of the meanings of the terms – then it cannot be the meaning of something’s being wrong that you or society are inclined to disapprove of it.

Another naturalistic approach to ethics that might fall under the terms of Moore’s naturalistic fallacy was that of R.B. Perry (1926). Like Dewey, he was influenced by pragmatism, in particular by William James, but, also by behaviourism. And in the iconoclastic spirit of early behaviourism he was much less cautious about making reductionist claims. He defined value as follows: “Any object, whatever it be, acquires value when any interest whatever it be, is taken in it; just as anything whatsoever becomes a target when anyone whatsoever aims at it.” (1926, section 49)

Reminiscent of Mill’s apparent confusion of desired and desirable, Perry identified having value with being valued rather than with being valuable, and then assimilated being valuable to that. It is precisely this sort of account that Moore was targeting. It is an open question whether there is any value in some thing or activity that a particular person is interested in. It is not an open question whether such a thing is valued.
But perhaps Perry could have conceded that at this stage of the story he was talking about something having value only in the sense of its being valued. The notion of right behaviour comes later when considering ‘overall value’. Overall value is determined by a harmony of interests. Like Dewey he considered the process of reflective agreement to determine ultimate value. And like Dewey, if this is taken to be the idea that ultimate value is a function of the process of reflective agreement then he was vulnerable after all to Moore’s naturalistic fallacy argument.

Certainly the development of this kind of approach into what became known as Ideal Observer theory seems to be so vulnerable. Roderick Firth (1952) defended this sort of analysis which construes “statements of the form ‘x is P,’ in which P is some particular ethical predicate, to be identical in meaning with statements of the form: ‘Any ideal observer would react to x in such and such a way under such and such conditions.’” (1952, 321) An ideal observer is omniscient, omnipercipient, dispassionate, disinterested and consistent (though in other ways perfectly normal!).

If the ideal observer and the appropriate conditions for the ideal observer can be characterized in naturalistic terms, then we have here a straightforwardly naturalistic definition of ethical terms. And Moore’s open question argument can be applied here. “An omniscient, omnipercipient, dispassionate … observer would react positively to this; but is it good?”

The reason that question seems open is that it seems to be possible to make the judgement about the reactions of an ideal observer without being thereby committed to have the same reactions oneself. “Why should I care how an ideal observer would react when working out how I should react? Even if an angel would judge that this is what should be done in this situation, I might say that’s irrelevant; I’m not an angel.”

A similar account to Firth’s was defended by Richard Brandt in Ethical Theory (1959). He argued that ethical terms usually corresponded to some particular sort of attitude. ‘Desirable’ corresponds to desire, ‘blameworthy’ corresponds to blame, ‘hateful’ corresponds to hate, ‘right’ corresponds to approval of the action, etc. So generally an ethical predicate means something like ‘worthy of such and such an attitude’. As a definition this is naturalistic only in the innocuous sense that Dewey and Aristotle were
naturalists: namely that ethical terms are anchored in natural aspects of our human condition. But it does not represent any sort of reductive definition, since the word ‘worthy’ is thoroughly ethical. The open question argument does not apply, since it makes no sense to ask why you should admire something that is worthy of admiration.

But then Brandt developed this in terms of a definition that has just enough substantial content to be subject to Moore’s argument. It is as follows:

The quasi-naturalist definition proposes that ‘x is E’ (where E is some ethical term) means the same as ‘The E-corresponding attitude … to x satisfies all the conditions that would be set, as a general policy, for the endorsement of attitudes governing or appraising choices or actions, by anyone who was intelligent and factually informed and had thought through the problems of the possible different general policies for the endorsement of such attitudes.’ (1959, 265-6)

Brandt had a view about what sort of method such an intelligent and factually informed person would adopt. He called it the Qualified Attitude Method and it amounts to something quite similar to what Firth required of the Ideal Observer. But mindful of Moore’s challenge, Brandt did not include it in his definition of ethical predicates. Still, what he did include is far from trivial. And it is certainly open to someone who is not particularly intelligent (whatever that is) or well informed to ask why they should approve of something just because it meets some condition laid down by someone who is.

Gilbert Harman’s picture of ethics, developed for example in his 1977 textbook, Morality, is subject to the same sort of worry. The picture is explicitly relativistic. “For the purposes of assigning truth conditions, a judgement of the form, it would be morally wrong of P to D, has to be understood as elliptical for a judgement of the form, in relation to moral framework M, it would be morally wrong of P to D. Similarly for other moral judgements.” (1977, 17)

This certainly does not look like a naturalistic definition of ‘morally wrong’, not least because the phrase ‘morally wrong’ figures on both sides of the explanation. But the problem arises when what it would be for something to be wrong in a moral framework
is understood, as Harman does understand it, as a claim about the attitudes and conventions of that moral framework. Something is taken to be wrong in a moral framework if an ideal reasoner adopting the conventions and basic attitudes of the moral framework would judge it to be wrong. Yet it seems a perfectly good open question to say: “I know that an ideal reasoner adopting my moral framework would judge this to be wrong, but is it really wrong?” To put it even more bluntly, a relativistic account like Harman’s may be criticised because there is always an open question as to whether one’s own moral framework is the right moral framework. And this question would be closed off if his relativism were correct.

In response to Harman among others there sprang up in the latter couple of decades of the century a group of moral realist philosophers often called Cornell Realists. These included Richard Boyd and Nicholas Sturgeon. Their view was that moral properties were natural kinds. Saying that something is good is like describing an animal as healthy. They claimed that although it is not possible to provide a definition, we can point to paradigms of good things or healthy things and conceive of the whole extension of the term as a natural extension of these cases. A similar suggestion was made later by Frank Jackson (1998) who argued that something is right if and only if it has whatever property plays the rightness role in mature folk psychology, and that this property will turn out to be expressed by an infinite disjunction of descriptive predicates.

Boyd argued that associated with a moral term is a cluster of mutually supporting natural properties – a ‘homeostatic’ cluster. Because the extension of the term is partly determined by this natural and causal idea of mutually supporting properties there is no objection in principle to moral properties being causally significant, and therefore real. According to Sturgeon the best explanation for someone’s judging that something is bad is that that thing is bad. Similarly, the best explanation for someone acting in a certain way might be that they are a good person. And this is what is required for realism.

The open question argument does challenge this approach in proportion to the degree to which the approach spells out what makes it the case that some natural cluster of

---

3 Boyd’s “How to be a moral realist” and Sturgeon’s “Moral Explanations” are both in Sayre-McCord (1988).
properties is associated with a moral term. The open question here is: ‘Why should I care about whether some action, situation or person belongs to such a cluster when deciding how to act? I can see that this thing belongs to such a cluster but is it really good?’

The Cornell realists have generally not accepted Hume’s internalist claim that moral judgments all by themselves can motivate people, and so may be unmoved by the openness of this question. Also they can appeal to the idea that the cluster of natural properties associated with a moral term is determined partly by a process of critical reflection – a process that precisely matches the process an individual must go through to work out what to do, how to be or what state to value. Like Dewey, these realists believe that there is a continuity between scientific method and ethical method. In both cases we look for explanations of our experiences in hypotheses about how things are. And in both cases we test these hypotheses, by considering new and untried situations. The real challenge then is to explain how what makes a set of natural properties cluster together in a self-supporting way coincides with the results of rational ethical reflection.

**Non-cognitivism**

Those philosophers inclined towards a strongly naturalistic approach but finding it unsustainable, perhaps because they accepted Hume’s internalism, found what they were looking for when in 1923 C.K. Ogden and I.A. Richards introduced the idea of emotivism. Ogden and Richards were linguistic philosophers concerned with how language sometimes symbolizes things and sometimes merely appears to symbolize things. In their book on language called *The Meaning of Meaning* (1923, 125), there is precisely one paragraph on ethical language.

“The peculiar ethical use of ‘good’ is, we suggest, a purely emotive one. When so used the word stands for nothing whatever, and has no symbolic function. Thus, when we use it in the sentence, ‘This is good,’ we merely refer to this, and the addition of ‘is good’ makes no difference whatever to our reference. When on the other hand, we say, ‘This is red,’ the addition of ‘is red’ to ‘this’ does symbolize an extension of our reference, namely, to some other red thing. But ‘is good’ has no comparable symbolic function; it serves only as an emotive sign
expressing our attitude to *this*, and perhaps similar attitudes in other persons, or inciting them to action of one kind or another.”

This idea chimed with the American pragmatists’ principle that the meaning of words was a function of how they were used and their rejection of the realist principle that this must always be understood in terms of how this use of words was related to some independently existing reality. The idea also coincided with the development and influence of logical positivism. Assuming that ethical language could not meet the conditions of properly scientific discourse, the hard-line positivist conclusion was that it was meaningless. But the positivists, especially the so-called Left Vienna School, were ethical thinkers like everyone else, and some significance had to be found for such thinking. The linguistic analysis of Ogden and Richards suggested what this significance might be. The function of ethical language was to express attitudes and persuade or incite people to act, but not to describe anything.

In other words the significance of moral judgements was *non-cognitive*. A moral judgement was not taken to be the application of a shared concept to a situation by the application of perception and reason. So the normal categories of truth and falsity were taken not to apply to moral judgements. Instead a moral judgement was taken to be the *expression* of a personal or interpersonal attitude, emotion, instruction or norm. As such it might be described as authentic or inauthentic, but not as true or false.

A.J. Ayer developed this idea in 1936 and Charles Stevenson the following year. Stevenson wrote: “Doubtless there is always some element of description in ethical judgements, but this is by no means all. Their major use is not to indicate facts, but to *create an influence*.” (1937, 18) C.D. Broad (1934) described it as the ‘interjectional’ theory. To say, “That act of self-sacrifice is good,” is to say something that functions precisely like “That is an act of self-sacrifice. Hurrah!” To say, “That deliberately misleading statement is bad,” is to say something that functions precisely like “That is a deliberately misleading statement. Blast!”

Non-cognitivism is a way of responding to Moore’s open question argument. In fact Darwall, *et al* (1992, 118) have described it as ‘the real historical beneficiary’ of that
argument. There are two possible reasons for this. On the one hand, you cannot accuse non-cognitivists of making the naturalistic fallacy, since they are denying that moral terms have meanings in the normal sense. Moral terms do not correspond with concepts that may be applied correctly or incorrectly to situations. So non-cognitivists are certainly not illegitimately defining moral concepts in terms of their extensions. Also the more general internalist worry with naturalistic accounts does not apply to non-cognitivists. There is no room for someone to say: “This is an action of giving money to charity – Hurrah! But so what?” Expressing the positive attitude to that action commits the speaker to accepting that the action is right – that they should do it.

However there is room for some sort of open question argument at an interpersonal level for a non-cognitivist. Suppose person A says to person B, “You should not smoke in this compartment.” Person B may reply: “I understand what you are saying – and see that you are expressing your disapproval of my smoking in this compartment, but why has that got any relevance to me? It is still an open question for me whether or not I should smoke.”

The problem for non-cognitivism is that when someone tells me what I should not do and I think I should do it I am disagreeing with them. But according to non-cognitivism I am merely not complying with their instruction or attitude. The basis of someone’s instruction to me or attitude about what I am to do is not automatically a basis for me to give myself the instruction or have the attitude, and so act in that way. And yet if I accept someone’s reasons for judging that I should not smoke in this compartment, it seems that I can’t deny that I have a reason for not smoking.

This is just to say that non-cognitivism, by eschewing objectivity, gives an account of moral judgements that falls short of what we might have thought was required by our actual practice of moral discourse. During the second half of the century non-cognitivism developed by putting back as much objectivity into the story as was possible while denying that moral judgements were simply descriptions of how things were. But however close to objectivity non-cognitivism got it never dealt with the problem of the interpersonal open question argument. For a non-cognitivist one person’s moral judgement being correct and appropriate provides no reason for another person to accept it; and this consequence just has to be swallowed.
So Richard Hare in 1952 argued for something he called ‘universal prescriptivism’. One minor departure from his emotivist predecessors was his insistence that the core function of moral judgements was to command or prescribe. He was less interested in the larger range of attitudes and feelings that they had taken to be expressed in moral discourse.

More significant was his introduction of universalizability to provide something like a descriptive element to his non-cognitivism. The idea was that the function of the claim “You should not smoke in this compartment,” was the same as that of the instruction “Do not smoke in this compartment!” But there was one extra aspect to the moral claim, namely its universalizability. The non-moral injunction does not commit you to the claim, “Do not smoke in any compartment exactly like this one in all morally relevant respects!” whereas the moral one does. Unlike mere instructions, the prescriptions expressed by moral judgements must be universalizable in that they commit one to the equivalent prescription in morally similar situations.

Hare’s theory generated a huge amount of argument. One problem was with the status of the universalizability requirement. It appears in the theory as a merely logical condition of a prescription counting as a moral prescription. But universalizability appears to be a substantive moral requirement of our moral discourse not a merely logical condition of that discourse counting as moral. A related problem was with the correct characterization of universalizability and in particular with the notion of morally relevant properties. It seems to be a notion that can only be characterized through moral reflection, which would threaten Hare’s account with a sort of circularity.

A further difficulty raised by Peter Geach (1958) is that the expressivist approach seems to break down when the moral part of a judgement is logically embedded in a larger judgement. We might accept that “Stealing is wrong,” is functionally equivalent to “Stealing – Blast!” But what about saying, “If stealing is wrong then taxation is immoral.” It seems to make no sense to identify this with, “If stealing – blast, then taxation – blast.” The “if then” structure does not operate on expressions of attitudes but only on descriptive components that have truth values.
One way to respond to the Geach problem would be to say that in this sort of context, the use of “if … then” is itself expressive. It expresses a commitment or an attitude of approval towards a certain transition of attitudes - the transition from disapproving of stealing to disapproving of taxation. Simon Blackburn (1984) explored this idea in his non-cognitivist account of Quasi-Realism. And Alan Gibbard (1990) did something similar by applying non-cognitivism to rationality itself.

**Moral Realism**

These developments of non-cognitivism in which our use of logical forms like “if … then” and our talk of normativity in general are taken to be expressive have taken non-cognitivism quite close to allowing talk of truth in moral judgements. An expressivist approach to truth itself, like that of Robert Brandom (1994) would in theory allow moral expressivism to be compatible with moral judgments being genuinely true and false.

The twentieth century debate about the nature of truth between correspondence theorists, deflationists, anti-realist assertibility theorists, etc (see chapter ??) has made the related issues of moral truth, moral realism and moral objectivism very complex. Anti-realists about truth would *ipso facto* be anti-realists about morality but might still accept that moral judgements were true or false if they satisfied the pragmatic constraints of a certain kind of discourse. One such constraint might be that moral attitudes must converge in response to their being open to higher-order attitudes about these very attitudes – i.e. something like critical reflection.

David Wiggins (1991) took the issue of whether moral judgements might be true to depend on an evaluation of whether moral discourse met such constraints – what he called “marks of truth”. And he argued that only some sorts of moral judgements met such constraints and only in a qualified way. In the process he made what he took to be a moral subjectivist position, inspired in particular by David Hume, as close to objectivism as he could.

A more extreme subjectivism was defended by John Mackie (1977) who argued that moral truth would require moral truth-makers – real values somehow out there in the world. And he argued that both metaphysically and epistemologically these real external values would have to be very queer things indeed – so queer that it would be
reasonable to assume that no such things existed. They were not discernible by normal perceptual means; so they would have to be discernible by some very queer extra faculty. They would have to both out there and in here simultaneously, since moral values were action-guiding, and that would make them pretty queer too.

Mackie was not content with the non-cognitivist response that our moral talk was not committed to the existence of such external standards but merely expressed our own internal standards and attitudes. He thought our moral talk was committed to the existence of these queer real external values and consequently just about every straightforward first-order moral judgement we make is strictly false. This was his error theory of ethics.

Some moral realists like the Cornell realists mentioned earlier tried to meet this challenge head on and argued that moral properties were real external causally active and explanatory features of the world. Others, like John McDowell (1985), denied that ethical standards were or needed to be blankly external, but neither were they mere expressions of actual attitudes. Helping himself to the analogy with secondary qualities – like colour or loudness - often employed by anti-realists, he argued that things were coloured red independently of any actual experience of this, but that to be red is to be such as (in certain cases) to look, precisely, red. So secondary quality judgements are grounded in the notion of human experience, but are still correct or incorrect independently of any particular experience or set of experiences – and so objective. Similarly, moral judgements are grounded in the attitudes generated by human sensibility, but are still objective.

Thomas Nagel, in a series of articles and books from “Subjective and Objective” (1979) onwards developed the idea of a certain kind of progression in ethics from subjective viewpoints to more objective detached ones in which one’s own interests and attitudes were removed from the viewpoint and treated as parts of the objective world along with everyone else’s interests and attitudes. For Nagel, our task in ethics is not to eliminate the subjective perspective, but to integrate it with more objective perspectives. Given this model, his way of understanding the question of moral realism then is whether from an objective viewpoint there would be any reasons to act.
His first book, *The Possibility of Altruism* (1970), employing a strongly Kantian approach, answered yes to this question. Altruism was taken to be a rational requirement for action and from this requirement moral principles could be derived. He began by pointing out that even self-interest theories of rationality take rationality itself to involve a substantial requirement – namely prudence. The very structure of means-ends rationality requires that the interests of oneself in the future are to be of concern.

In the case of altruism the argument is that if you have a reason to further your own interests or meet your own needs you are committed to thinking of such needs and interests as providing reasons objectively, and thus as providing reasons for other people too. This then commits you to thinking of other people’s needs and interests as providing you with reason to act. In this way Nagel mirrored Moore’s (1903) argument against ethical egoism.

Christine Korsgaard (1996) has developed this Kantian project, and at the same time attempted to make a synthesis between it and Aristotle’s approach to ethical behaviour. She claims that by reflecting on our nature as agents (on our ‘practical identity’) we find that consciousness of our own humanity gives us a kind of authority for action; and “it is this authority that gives normativity to moral claims.” (1996, 20)

**Intuitionism**

Mackie’s attack on external ethical values as having to be queer both metaphysically and epistemologically was in part an attack on G. E. Moore’s conception of the good as a simple non-natural property knowable through intuition. Moore’s own approach had a huge influence in the first half of the twentieth century, although it was itself strongly influenced by a work of the nineteenth century, Henry Sidgwick’s *The Methods of Ethics* (1874), whose seventh edition was published in 1907.

While seeing no alternative to utilitarianism as an account of what actions are right, Sidgwick and Moore were fiercely critical of Mill’s half-hearted attempt to ground the principle of utilitarianism in definitions and proofs. Instead they talked of intuitive knowledge. As I will try to show here, this intuitionistic approach developed almost
seamlessly into the much more Aristotelian approaches of late twentieth century sensibility theories of ethics.

But Sidgwick explicitly rejected the Aristotelian idea that decisions as to how to act in particular situations were to be arrived at through some kind of intuitive conception of the merits of the case. He described this as the phase of perceptual intuitionism, and thought of it as a primitive stage in the process of moral development. The next stage is to ground these judgements in general principles and rules. But when you have no argument for these principles and rules of conduct, but merely intuit them as correct, you are at what he calls the phase of dogmatic intuitionism. You can do better than this by grounding your choice of principles and rules in some fundamental philosophical principles about the nature of ethics. But these philosophical principles cannot be grounded in anything more fundamental. The process of providing ever deeper justifications must stop, not as Mill suggested in some definition of ‘desirable’, but in fundamental unargued intuitions. This is the phase of philosophical intuitionism, and we must rest there.

However Sidgwick’s construction of a sequence of phases in the rational development of ethics positively invites a further extension of the sequence. Why should the correct choice of fundamental philosophical principles when faced with a decision between incompatible ways of grounding our rules of conduct not be determined by some further justification rather than by intuition? This question seems particularly pressing since what he takes to be the basic philosophical principle underlying ethics now seems so arbitrary – namely that the right action is the one that will maximise the good.

Moore criticised Sidgwick’s claim that human happiness is the only thing good in itself, arguing, perhaps rather strangely, that faced with an alternative between two worlds, one beautiful and one ugly, neither of which contained any people or other creatures capable of judgement, it would by rational to try to bring about the beautiful one (1903, 83 ff.). But Moore accepted Sidgwick’s claim that we can know through intuition the general truth that the right action is the one that will produce the greatest possible amount of good. And in this respect his theory of what actions are right (as opposed to what things are good) was criticised by W.D. Ross (1930). Applying our moral intuition sensibly, we can still say, “I can see that doing this would maximise the
amount of good, but is it right for me to do it?” For example, suppose that our intuition is that in some situation where utilitarianism is in conflict with a conception of justice based on rights and obligation justice should prevail. This would mean that utilitarianism itself was not an intuitive truth.

Ross, an Oxford moral philosopher influenced in about equal parts by Aristotle and Kant, (1930, 17) argued that we do not always look to the future, but often to the past, when determining how to behave. So Ross, following H.A. Prichard (1912) and in parallel with C. D. Broad (1930), rejected Sidgwick’s and Moore’s idea that the issue of how to act – what is right – depended on some grand intuitively knowable principle linking right and good. Just as Moore rejected any grand \textit{a priori} principle concerning what is good, Ross rejected any such overarching principles concerning what is right, and argued that knowledge of what is right is based on more piecemeal and particular considerations.

Ross claimed that there is a set of morally relevant considerations that give us reason to act – what he called \textit{prima facie} duties – some of which are consequential and others of which are not. Broadly they come under the following headings: fidelity; reparation; gratitude; justice; beneficence; self-improvement; and non-malificence. As with Sidgwick’s and Moore’s grand consequentialist principle, our knowledge of these \textit{prima facie} duties is intuitive. Broad said similar things, talking about considerations that made actions \textit{fitting} or \textit{unfitting}.

Since there is a plurality of \textit{prima facie} duties they will often conflict. Whereas we can know intuitively what are the \textit{prima facie} duties, we do not always know how to weigh them, whether we have missed any in the particular situation and, with respect to the consequentialist ones, what the consequences are. So we can only come up with more or less probable judgements about the rightness or wrongness of particular actions.

Ross’s approach to ethics generally was echoed much later, when medical ethics started to develop. An influential book by Beauchamp and Childress (1979) argued that there were four basic principles of medical ethics: autonomy, non-malificence, beneficence, and justice. Although looking like a sort of ragbag approach to ethics, this has proved to be a remarkably robust tool in this branch of practical ethics.
The faculty of intuition sounds like a fig leaf for mere opinion or prejudice. And it has been a common criticism of intuitionism that talk of intuition just waves a hand at the epistemological problem rather than providing an answer. The criticism was particularly acute when targeting Moore’s idea that we could be intuitively aware of simple non-natural properties. But with Ross’s idea that good and right are complex things, it is not so clear that we have to posit some mysterious faculty to explain our intuitive knowledge of them.

The key thing about intuitive knowledge is that it is non-inferential. The intuitionists were keen to deny that our knowledge of what is good (and of what is right for those, unlike Moore, who thought that ‘right’ was not definable either) is derived through inference from some more basic principle or definition. But it does not follow that perception and reason do not go to make up intuition. If a tree surgeon looks at a tree and judges that it is unhealthy they may not be inferring this from some principle that says that whenever a tree has a certain property it is unhealthy. They are able to recognise that it is unhealthy because they have refined their capacity to apply this concept. There is no need to posit a special faculty, just a special conceptual or recognitional skill.

Aristotle’s ethical insights were employed to develop this idea in intuitionism in several ways simultaneously, very often by Oxford philosophers. One Aristotelian development of intuitionism was particularism – the idea that moral judgements did not need to be grounded by general principles. Ross made some use of general principles with his list of *prima facie* duties, but these were not supposed to determine moral judgements even though they provided some grounds for them. Later particularists, like John McDowell (1979) and Jonathan Dancy (1983), rejected even this role for principles, arguing that sensitivity to morally relevant properties need not be mediated by recognition of the application of any universal principles.

Such sensitivity must of course be learnt and developed. Moral sensibility must be refined. This is another feature of Aristotle’s approach that fed into later twentieth century successors to intuitionism. The approach, sometimes called sensibility theory, accepted Aristotle’s principle that moral judgement depended on the exercise of a
sensibility constituted from properly refined and rationally moderated emotions. It was still a conceptual exercise since these emotions were subject to justification, but one grounded in human nature.

G. H. von Wright (1963, 171) developed a conception of practical reasoning rooted in the knitting together of “wanting an end, understanding a necessity, and setting oneself to act.” Martha Nussbaum (1978, 178) described the appeal of explanations of action in terms of such a notion of sensibility as lying “in their ability to link an agent’s desires and his perceptions of how things are in the world around him, his subjective motivation and the objective limitations of his situation (as he sees them).” David Wiggins (1987, essays 5 and 6) found this idea in Hume and reflected in his non-cognitivist successors, as well as in Aristotle’s notion of aesthesis, which Wiggins translated as situational appreciation.

In this tradition the notion of intuition amounts to something like that of seeing things properly. It is bound up with having the capacity to feel things properly. Emotions are rationally assessable. And, as John McDowell (1979) has argued, the ethical concepts that one applies on the basis of this knitting together of reason, perception and emotion are irreducible to concepts that one can grasp fully from a non-emotional perspective. Nevertheless, moral judgement is taken to be an exercise in concept application, just like any other form of judgement that aims to be true. It is just that the ability to apply these concepts properly depends on having the sensibility that goes with properly realized and moderated emotions – i.e. it depends on being virtuous.

This is why Moore’s open question argument has no force against a thorough-going Aristotelian sensibility theory. The initial worry might be that it makes sense for a non-virtuous person to say: “I can see that that is what a virtuous person would do; but not being perfectly virtuous myself, why should I care?” Identifying the right act with the act of a good person appears to risk alienating normal people from morality. But McDowell’s point is that being virtuous enables you to tell what is the right act. He is not simply recommending the following decision procedure for the less than fully virtuous person: work out what the virtuous person would do and act accordingly. For this decision procedure is not fully available to the non-virtuous person. There may be good reasons to emulate people who seem virtuous to you in order to acquire virtue, but
this is a much weaker claim than the putative decision procedure. McDowell’s point, echoing Moore (but more obviously Aristotle) is that knowledge of what is good or right is not available from outside the practice of moral thinking, and such a moral practice requires the properly moderated and modulated emotional sensibility that goes with being virtuous.

This idea served feminist approaches to ethical knowledge. The feminist psychologist Carol Gilligan (1982) developed a distinction between two ways of approaching a moral problem. One way, which might be associated with a man’s approach was to look for a solution to the problem as stated. Principles are brought to bear and argued about, and eventually one side or other of the argument is assumed to win. The language is adversarial.

The other way, which might be associated with women, is to work around the problem; not to accept the problem as stated but to redefine it if possible. No situation comes fully conceptualised. Any description of a moral problem can be improved upon until it looks much less like a moral problem as such. By talking through the situation with lateral thinking, it might be possible to avoid a clash of principles, but to reach an acceptable course of action through consensus. On the one hand is defined the ethics of justice and duty. On the other hand is the ethics of care. Aristotelian sensibility theory provided an intellectual arena for the ethics of care.4

According to Bernard Williams (1985), the various ethical concepts are more or less ‘thick’ ranging from concepts like ‘gratitude’, ‘brutality’, and ‘bravery’, which are ineliminably ethical but have a lot of descriptive content to thin concepts like ‘good’ and ‘right’. No attempt to give an account of the application of thin concepts can work without recourse to thick ones, but the application of thick ones is itself a messy business. In the same way Philippa Foot (1958) had argued that evaluative concepts like ‘rude’ have clear conventionally accepted criteria for application and still express the attitude of disapproval. The only way to avoid being committed to such an attitude when faced with these clear criteria is to refuse to use the concept of rudeness.

4 To be stuck with the ethics of care has not suited all feminist thinkers. Sandra Bartky (1990) has warned of the consequences for disempowerment.
As Iris Murdoch (1970, 34) wrote, “Moral language which relates to a reality infinitely more complex and various than that of science is often unavoidably idiosyncratic and inaccessible.” Murdoch’s work in the 1960s culminating in her 1970 book, The Sovereignty of Good, was highly influential in the development of sensibility theory in Oxford. More influenced by Plato than by Aristotle and more concerned to respond to French existentialist philosophy than most of her colleagues were, she brought to bear Simone Weil’s (1959) notion of ‘attention’ to capture the idea of a loving and humble gaze directed upon particular individuals and situations. The central idea was that there is no way to fix a description of a particular person and their situation or even fix the concepts that apply to them, but that the process of trying to perfect our ways of seeing things is essentially the moral project – a project in which art and literature are more useful than more theoretical pursuits.

One of the important contributions of Murdoch’s work was to make a bridge between the work on ethics in the analytical Anglo-American tradition and the work being done in an existentialist or phenomenologist mode on the continent of Europe. For example, Anglo-American sensibility theory bears a striking resemblance to the idea of Martin Buber (1937) that ethical understanding requires adopting a non-objective perspective towards others – what he called an ‘I-thou’ relationship.

In one respect this was to treat people in a Kantian way as always ends in themselves rather than as merely means to achieving ends. But Buber rejected Kant’s conception of the moral law, arguing, in the spirit of existentialism (and indeed Aristotelian particularism), for what became known as ‘situation ethics’. A concrete situation of infinite particularity is presented to us, and what constitutes our responsibility is to answer to this situation in the fullness of one’s being. Somehow this also constitutes our relationship with God. This idea was developed by Emmanuel Levinas in his Totality and Infinity (1961). There he argued that ethics involves an encounter with the unknowable Other.

The role of emotions in our moral sensibility was more explicitly brought out in the work of Max Scheler (1954) – defining what became known as ‘axiology’ or the philosophy of values. Values, for Scheler, were phenomenologically disclosed in emotional experience. His conception of an emotion was of an experience with both a
subjective and an objective aspect – the objective aspect revealing real values. To this extent our conception of the evaluative world was partly constituted by our nature as feeling creatures, while our nature as feeling creatures was partly constituted by our capacity to perceive values. This virtuous Aristotelian circle of values and emotions was precisely what McDowell and other sensibility theorists were so keen to defend.

In 1958 Elizabeth Anscombe mounted an attack on all moral systems based on general principles, whether consequentialist or Kantian. In the same way as Nietzsche had attacked morality generally, Anscombe argued that there was no place in current ethical thinking for an authority backing up these moral obligations. The only role left for sensible moral discourse was in understanding human virtue; and this would need to wait until moral psychology was more developed.

This paper is often taken to herald what is known as Virtue Theory - the idea that actions are right or wrong in virtue of whether or not they issue naturally from good or virtuous character dispositions. Virtue Theory is an aspect of Aristotelian sensibility theory, since according to sensibility theory, having good character dispositions is part of having a properly moral sensibility, which is what enables one to tell what is the right thing to do in any particular situation. And knowing what is right will naturally lead to acting well.

So sensibility theory leads to the slogan that the right act is the one that the good person will perform, which is Virtue Theory. But it gets to this slogan via the Socratic claim that virtue enables you to know what is right. And this claim about knowledge is not always taken to be part of Virtue Theory. For example Alasdair MacIntyre (1981) did not identify virtue and knowledge, but argued that we can gain insight into a culture’s moral practice by investigating its conception of moral virtues. His approach, in the spirit of Virtue Theory was to reject the search for a grounding of our concept of virtues in anything else. But, in the spirit of the postmodernism of the 1960s and 70s he went on to make the further move that there was no room for rational argument in favour of one or other conception of virtue. This move was certainly not warranted by Aristotelian sensibility Theory.
Bernard Williams (1973) while not endorsing anything like a systematic virtue theory, was in the forefront of the argument for so-called agent-centred ethics. He argued against utilitarian and Kantian approaches by claiming that they failed to acknowledge the particular moral viewpoint of the individual agent. Moral thinking must be relevant to the moral agent; it must be action-guiding. And as such it must reflect the agent’s interests and concerns. It must respect their integrity as a moral agent.  

For Williams, this position was associated with a view about moral reasons and indeed about practical reasons generally. In his paper, “Internal and external reasons” (1981, chapter 8) he argued that anything that counts as a reason for someone to do something must be grounded in their motivational set of desires and other attitudes. Something just does not count as a reason for someone to act if it cannot be derived from this motivational set.

The significance of this idea depends to a large extent on what resources can be appealed to in the process of deriving reasons from a motivational set. Although Williams would have rejected the possibility, it is compatible with his view of internal reasons that some moral reasons may apply to everyone and still count as internal reasons inasmuch as they could be derived from any motivational starting point whatsoever once it was properly worked out. In any case the paper represented the start of a resurgence of interest in the nature of practical reasons. For example, Michael Smith (1994) defended a Humean conception of what he called ‘motivational reasons’ in which such reasons were determined by the agent’s beliefs and desires while rejecting Williams’s claim that all reasons for action (including what Smith called ‘normative reasons’) were rooted in the agent’s beliefs and desires. On the other hand Jonathan Dancy (2000) argued against a psychological conception of motivating reasons as well; only in odd cases do one’s beliefs and desires count as reasons for one to act. The issue is complicated however, with various important distinctions – e.g. between reasons for acting and reasons why one acts – still not having been properly worked out by the end of the century.

---

5 This sort of criticism of utilitarianism has itself been criticised by Shelley Kagan (1991), who has maintained that moral requirements may diverge from intuitively acceptable requirements, in which case, as moral agents, we should aim to change ourselves so that what seems acceptable lines up better with what is morally required.
Agency, free will and responsibility

Underlying the central issue of how morality is related to the will is the question of what agency is. Hume identified agency with the causal role of a subject’s psychological states – their beliefs and desires – while Kant identified agency with the role of practical reason. These two approaches continued to characterize work on the philosophy of agency through the century, with the high point perhaps being Donald Davidson’s (1963) article bringing Hume and Kant together by characterizing the primary reason in the light of which an action is intelligible as being constituted by a belief/desire pair, and then insisting that explaining an action in terms of reasons is simultaneously to make it rationally intelligible and to cite causes that explain why it happened. In claiming that reasons were simultaneously causes he was replying to the influential non-causal approaches then prevalent deriving from Gilbert Ryle (1949), Elizabeth Anscombe (1957) and Alfred Melden (1958).

There was still an issue as to whether agency in this sense was sufficient for moral responsibility. Many philosophers thought that it made sense to say that someone might act but not of their own free will and so not be morally responsible for doing so. Two principles in particular were often suggested as providing further requirements of free and therefore responsible agency. The first was the causal origination principle - that to be responsible for one thing you must also be responsible for what caused it. Philosophers from Paul Edwards (1958) to Galen Strawson (1986) have defended this principle and at the same time argued that we are not responsible for the ultimate causes of our achievements and so not responsible for these achievements either.

The other principle, often called the principle of alternative possibilities, is that to be acting freely and responsibly it must be possible that one acted otherwise. This is an attempt to capture the general ethical principle that ‘ought’ implies ‘can’. If there is no alternative possibility then it can never be right to say that you ought to have done otherwise and so it is never right to blame you for doing what you did do. This principle was used by for example van Inwagen (1975) to argue for the incompatibility of free will and determinism.

What bothered these twentieth century philosophers as much as it had bothered earlier philosophers was whether there was space for the idea of moral responsibility in the
naturalistic scientific world picture where every event was either determined by previous events according to the laws of physics or was the realization of some quantum probability function. The “hard” determinists thought this was a real problem. The “soft” determinists or “compatibilists” thought that free will and responsibility were not threatened by any of this. And the libertarians rejected that particular aspect of the scientific world view that seemed to threaten free will, namely determinism.

One important version of this rejection of determinism was the idea of agent causation, developed in different ways by Richard Taylor (1966) and Roderick Chisholm (1966). Taylor was concerned that an account of agency that worked with the idea of mental events causing behaviour missed out the central role of the agent themselves in agency – a role that could only be accommodated by assuming that the idea of an agent making something happen was irreducible to that of states of the agent making things happen. Chisholm’s concern was with free will, arguing that the possibility of free will required that some events were not caused by other events but only by the agent. While this came at too high a metaphysical price for most philosophers since 1966, there were some notable exceptions later in the century – for example Tim O’Connor (1995).

Many compatibilists – e.g. Daniel Dennett (1984, chapter 6) - have remarked that the question of whether you could or could not have done otherwise in some situation depends entirely on how the notion of possibility here is being understood. There is no single answer to the question of whether you could have jumped four foot in the air. It depends on what is taken as given and what is allowed to vary when considering the possibility. So it is argued that a fallacy of equivocation is made in concluding from the deterministic (or fatalistic for that matter) claim that things could not have been otherwise to the incompatibilist conclusion that I could not have done otherwise.

The argument was dealt a further body blow by Harry Frankfurt in (1969) who argued that it is in any case not relevant to a consideration of free will or responsibility whether or not you could have done otherwise. If there was someone standing behind you ready to intervene (perhaps by manipulating your brain) if you were on the verge of deciding not to do some vile act that you were contemplating, then you are still responsible for that vile act when they do not have to intervene, even though you could not have done otherwise.
Peter Strawson’s (1962) article, “Freedom and Resentment”, presented a more radical rejection of the free will/determinism dialectic. He argued that whatever we think of grand metaphysical arguments against a certain sort of liberty this has no relevance to our attributions of moral responsibility. These grand arguments depend on adopting an ‘objective’ point of view when considering someone’s agency – i.e. treating them as an object rather than as one of us. The perspective from which moral responsibility is attributed is one in which our conception of someone’s agency is partly constituted by our own reactive attitudes to them – or the reactive attitudes that would be appropriate towards them. These are attitudes of resentment, gratitude, etc. The grand metaphysical arguments have no role in this perspective, nor do they undermine it.

In the spirit of the Aristotle renaissance in Oxford at the time, Strawson developed an account of how we attribute responsibility to an agent in terms of a properly tuned reactive sensibility towards them. He argued that we build up our ability to attribute responsibility from a starting point of being able to feel resentment and gratitude to others for the respect or otherwise that they show to us. We then develop the ability to feel vicarious resentment and gratitude on behalf of others, and then to say when such resentment and gratitude would be appropriate. In some circumstances it might not be appropriate to resent someone who acted freely and wrongly if they were in a very difficult situation, and in some circumstances it might be appropriate to resent someone for the lack of respect they showed in their behaviour and general attitudes, whatever their status as a ‘free’ agent. At this point when our reactive attitudes are rationally moderated we have developed a concept of objective responsibility.

In a more Kantian vein, Charles Taylor (1976) developed the view that determination of the will by reason constitutes free agency, but added the Hegelian twist that rationality was essentially dynamic. So freedom consisted in the capacity when faced with incompatible courses of action recommended by different evaluations to reconceptualize one’s situation and develop a further evaluation that incorporates the earlier ones and avoids the contradiction. This act of reconceptualization is an act of self-definition very much in the spirit of the existentialist identification of freedom and self-definition.
Taylor disagreed with Sartre (1948) however by claiming that an act of self-definition, although it is not itself derived from any deeper more basic rational evaluation is not somehow outside of rationality. The new conception of oneself and one's place in the world is straightaway subject to rational assessment and the possibility of revision even though there may be no competing evaluation that requires such an assessment straightaway. And in this sense such self-definition is sensitive to rationality.

For the existentialists, self-definition was a radical choice – not only guided by no more basic conception of oneself – but also insensitive to rationality as a whole. It was a blind leap. In virtue of being unconstrained even by rationality as a whole this possibility was supposed to capture freedom in the very strongest sense. Value was there to be created by such a leap; in this respect twentieth century existentialists set themselves apart from Nietzschean scepticism. But value was not external to such self-definition and could not guide it.

Starting from this idea of radical choice and of authentic moral thinking emerging from such unconstrained choice, it is not surprising that the existentialists had relatively little systematic to say about ethics. An exception was Simone de Beauvoir (1948), who argued that freedom was an end in itself and that dishonesty, inauthenticity and tyranny were all failures in this regard. But no definitive answers were provided by these generalities. Every instance of something having significance has to be created freely, and is always open to being recreated differently. And every realization of freedom in some act of self-definition, by fixing meaning at the same time constrains freedom. This is the tension that she called ambiguity.

**Utilitarianism**

The general principle that one should promote the good by one’s actions guided much ethical thinking in the twentieth century. This principle by itself might be innocuous, but when combined with the assumption that the good to be promoted can be understood independently of the particular agent’s concerns, obligations and entitlements, we get the theory known as consequentialism. If the good is treated as
happiness in some sense, then we have utilitarianism; and if the good is taken to consist of other things too, like beauty, knowledge, etc, then we have what Hastings Rashdall (1907) dubbed ‘ideal utilitarianism’. Sidgwick was a utilitarian – indeed a ‘hedonistic’ utilitarian since his conception of happiness was a hedonistic one in which happiness was understood in terms of pleasure and enjoyment. Moore was an ideal utilitarian, since he included other things than just happiness in his conception of a good state – for example beauty. More recently, James Griffin (1986, 67) presented a kind of ideal utilitarianism in which the central prudential values were accomplishment, fulfilling requirements of human existence and flourishing, understanding, enjoyment and personal relations.

Utilitarianism based on a hedonistic conception of happiness as pleasure or enjoyment and indeed any conception of happiness as an internal mental state is often thought to have fallen foul of Robert Nozick’s (1974) example of the experience machine. Nozick asks you to think of the things that you would most value achieving and experiencing in life and offers you the imagined chance to go into a machine that would simulate in complete detail the experience of achieving such things. If you valued writing a great book, you could go into the machine and become convinced that you really did write such a book and feel all the satisfaction that would come from that. Would you choose to go into the machine? If not, then what you take to matter is not an internal mental state of happiness but actual achievements.

The most influential departure from a hedonistic conception of happiness was a preference satisfaction conception. Although there were hints of this sort of approach in Mill’s Utilitarianism, it received a huge boost, when Frank Ramsey (1931) developed a behaviourist calculus of preferences. He showed that it was possible to construct a measure of utility associated with different possible outcomes just by determining how a subject would choose between different outcomes. He also showed that it was possible to construct a measure of their degree of belief in certain possible outcomes by determining how they would bet on these outcomes.

Ramsey’s aim was to develop a subjective approach to probability. But his calculus quickly became the basis of a self-interest conception of rationality – rational choice theory. This was developed into game theory in 1944 by von Neumann and
Morgenstern. The idea was that the rational thing to do is to satisfy your preferences optimally; and Ramsey had provided the basis of an axiomatisation of what this involved. John Harsanyi (1982) worked this up into what he called ‘preference utilitarianism’.

It is often thought that this approach as an account of rationality is based on a fairly simple fallacy. It is trivially true that you only ever do what you want. So it is supposed to follow that you only ever aim for what you want. You only do what you prefer to do; so you should do what you prefer to do. The desired becomes the desirable, as in Mill. The valued becomes the valuable as in Perry. If satisfaction of your preferences is taken to be the fundamental axiomatic principle of practical rationality, then these preferences are themselves taken to be determined outside of practical rationality. They are not the results of moral evaluation and consideration of others, but are given psychologically. We seem to have moved from the truism that you only do what you want to do or prefer to do to the extraordinary claim that rationality consists entirely in responding effectively to your preferences.

But as a way to approximate a measure of happiness, preference theory was a shot in the arm for utilitarianism. It turned out that utility can be measured after all. In social policy, cost-benefit analysis and welfare economics were born. Civil servants applied utility calculations based on experimental research into people’s preferences to establish the degree to which large scale projects would lead to an overall preferred outcome by the population. The practical significance of this branch of ethics was immense.

A central issue for utilitarianism, though one that at the same time spread into other areas of moral philosophy in the twentieth century – especially theories of rights - was the question of which beings were to be included in calculations of overall happiness. Whether one is a utilitarian or not, morality is concerned with respecting the interests of others; but which others should one be concerned with? Peter Singer’s (1975) Animal Liberation made the case on behalf of utilitarianism that all (and only) sentient beings counted in any estimate of overall happiness. This has led him to bite the bullet in many controversial ethical debates concerning animal rights, abortion and euthanasia. Abstracting from other factors that Singer concedes are bound to muddy the question in
practice, he argued that if a newborn child has less capacity for feeling pain than a cow we should be more concerned with the well-being of the cow.

Derek Parfit has raised the same question about future generations. In his (1977) *Reasons and Persons* he argued that as yet unborn people count alongside existing beings in one’s moral concern for the interests of others. He argued against a strict conception of personal identity; so even one’s concern with one’s own interests in the future is a concern with a distinct though psychologically connected being – one that does not yet exist. Along with the problems that the interests of future generation raise for utilitarianism, issues are raised in practical ethics concerning how we should treat the environment and whether we should apply genetic filtering to human embryos to ensure that children are born with the best possible set of genetic resources.6

**Rule-based systems**

The law was another area of public morality to which utilitarianism had been fruitfully applied since the time of the nineteenth century philosopher of law, John Austin. The structure and details of the judicial system could be put into a utility calculation and the system of laws that maximised overall happiness would be recommended. Indeed having any such system at all could be justified in this way. But what could not be justified in this way was an individual’s decision to be bound by the law - or indeed any self-constructed law either – when conforming with the law did not advance happiness.

If a judge knows that as a good utilitarian they should abandon proper procedure and summarily dismiss the case of someone they know to be a murderer in order to avoid riots and massive loss of life, then even if the law itself is justified on utilitarian grounds, utilitarianism dictates that the judge should ignore the law and release the murderer. Anti-utilitarians have argued all through the century that this sort of case shows that as a basis for individual decision utilitarianism leads to patently unethical results and fails to make proper space for the ideas of rights and justice.

6 See e.g. Steiner (1994).
The difficulty is that the utilitarian justification of following a rule is contingent on expected consequences, whereas our intuitions about rights and justice suggest that rules are obligatory in a way that is at least partially independent of the consequences of particular instances of following or not following them. This raises the question of whether it might be possible to provide a utilitarian justification of the claim that one’s action should be in accordance with the best set of universal rules or principles.

The idea would be that although following a particular rule on a particular occasion might not be justified on utilitarian grounds, being a rule-follower might be justified on such grounds, and the decision on a particular occasion might then be fixed by this prior decision. So before the judge calculates whether letting the murderer go free has better consequences than convicting him according to the principles of justice, he or she should decide the prior issue of whether or not to be principled.

The difficulty for this sort of utilitarianism was that only the weakest conception of being principled could possibly be justified on utilitarian grounds. It is useful to have rules of thumb to apply to situations in which the expected cost in happiness from spending a long time trying to make an accurate utilitarian calculation outweighs the expected gain from getting the calculation right. But this will not help the judge in the example just given. We can assume that he or she knows straightaway that releasing the murderer will result in more happiness than following through with justice. So whatever rule of thumb they are employing they should still in this situation do the unprincipled thing.

Another way to justify being principled is to justify having a principled character. One of the things a good utilitarian should do is act to develop in themselves the character that will lead to the best consequences. It may be that having a horror of injustice is a good thing to have in one’s character from a utilitarian point of view (though this would need quite a good argument). So the utilitarian judge would in this case feel a horror in letting the murderer go. Robert Adams (1976) developed something like this view in defending ‘motive-utilitarianism’.

But the utilitarian judge should still let the murderer go despite feeling horror. Richard Hare (1981) developed the distinction between intuitive level thinking and critical level
thinking to make just this point. Rules of thumb, deeply felt principles and character
dispositions might be developed according to utilitarian principles, and these might lead
one to be more principled than utilitarianism on an act by act basis (act utilitarianism)
would do. Acting according to such rules or character dispositions is working at the
intuitive level. But the good utilitarian should have the capacity to go against these
intuitive level recommendations and when the need arises apply a critical utilitarian
approach to whether or not to follow such rules or character dispositions in a particular
situation. The judge faced with the possibility of riots if he or she does not release the
murderer would have to apply critical level thinking to the case, and would have to
choose the unfair course of action as a result.

The difficulty for this sort of utilitarian justification of rules can be put in a different
way. If the principles or rules are supposed to apply across the board and not be
trumped and undermined by further critical thinking, then the rules would have to be
highly qualified. For example, the judge would have to construct a rule that went
something like this: “Follow the rules of the system with complete impartiality unless
not doing so would have a clear benefit for overall happiness.” Principles play no
significant part in this system.

Faced with the impossibility of constructing a utilitarian justification of following rules
in morality, R. F. Harrod (1936) employed something more like a Kantian justification
of following rules, when he introduced the idea of rule-utilitarianism. His idea was to
combine utilitarianism with the so-called generalization principle that in itself has
nothing to do with utilitarianism.

According to the generalization principle a rule is good if it would be good if everyone
followed it. If we can appeal to a Kantian justification of basing moral action on
principles and also appeal to something like Kant’s first formulation of the categorical
imperative – act only on that maxim that you would at the same time wish to be
universal moral law – then utilitarianism might still have a role in determining which
would be the best set of universal moral laws. In other words, utilitarianism was taken

to plug what was perceived to be a serious gap in the Kantian system – the transition from the formulation of the categorical imperative to substantial moral principles.7

As David Lyons (1965, 136) formulated it in an influential criticism of the different possible forms of rule-utilitarianism: “An act is right if, and only if, it conforms to a set of rules general acceptance of which would maximise utility.”

Many utilitarians have taken this proposal to be flawed as it simply rejects the basic utilitarian claim that when deciding how to act you should aim at achieving the best consequences. For example J. J. C. Smart (1956) maintained a sustained defence of act-utilitarianism in the face of this hybrid theory. One of his arguments was that even this more Kantian rule-utilitarianism must collapse into act-utilitarianism. Wouldn’t the world be as good as it could be if everyone simply followed the act-utilitarian principle that they should do their best to maximise overall happiness? Against the response that no one would trust anyone in such a world one might reply that in a world where you could rely on everyone being a good act utilitarian there would be no need for institutions of trust. As Lyons (1965) argued there are ways of formulating rule-utilitarianism that do not collapse into act-utilitarianism, but these are not always the most plausible theories in themselves.

A parallel debate was carried on by those moral philosophers who, in the spirit of Thomas Hobbes, were trying to base an adherence to moral principles on self-interest instead of overall happiness. Many philosophers in this tradition were persuaded by the argument that it is in all of our interests to have a system of rules that everyone sticks to. One’s world is safer; interpersonal relationships are stronger; generally one will be happier and more able to achieve one’s goals in a world where everyone is principled. So morality seems to be justified on the grounds of self interest.

Kurt Baier (1958) was a key figure in the development of this sort of approach. He argued that morality involved a special sort of reason – one that took account of the interests of others. In this he was following Stephen Toulmin (1950). But Toulmin did not attempt to ground moral practices in self interest but in their conduciveness to an

7 As Urmson (1953) argued there are signs of this sort of approach in Mill’s own utilitarianism. Other significant rule utilitarians include Mabbot (1953) and Brandt (1963).
overall harmony of interests. For Baier, moral reasons overrode those of simple self interest, but the fact that they had this priority was itself supposed to be grounded in considerations of self interest.

The challenge for this sort of approach was the Prisoners’ Dilemma. While it may be the case that we would all be better off if we all followed rules rather than all seeking our own self interest in an unprincipled way, it would still be in any particular individual’s interest to cheat when presented with a case where being principled would go against their interest. So every individual would cheat in their own interests, and as a result everyone would be worse off than they would have been if everyone had been principled.

If the only principle determining rational action is self interest, it looks as if people will not work within a system of moral principles at all, but will be bound to collapse into unprincipled behaviour which paradoxically leaves them worse off than in the moral system. So some further Kantian constraint on rationality has to be found if anything recognizable as a moral perspective can be grounded.

In this spirit David Gauthier (1967, 1986) argued that rationality essentially involved commitment. Commitment was part of the very idea of having intentions. Out of this he tried to argue that the self-interested decision would be to commit oneself to morality, where such a commitment would carry one through prisoners’ dilemmas. Alan Gewirth (1978) introduced the idea that it was in one’s interests to have generic rights respected, and argued that it would be irrational to insist on one’s own rights without respecting the rights of others.

The idea that rights were fundamental to ethical thinking was one that characterized much moral philosophy of the twentieth century. For example Judith Thomson (1976), developing Philippa Foot’s (1967) discussion, used the idea of rights to explain why it is morally permissible to turn a trolley from a track where it was going to kill five people onto a track where it will kill just one, whereas it is not morally permissible for a surgeon to kill one person in order to harvest his/her organs to save five others. She also worked out an influential argument in the abortion debate (1971), which appealed to the mother’s right to have a say in who was allowed to feed off her body.
Joseph Raz (1984, 195) developed a powerful theoretical proposal for linking rights with duties and duties with interests. “x has a right if and only if x can have rights, and other things being equal, an aspect of x’s well-being (his interest) is a sufficient reason for holding some other person(s) to be under a duty.” Raz’s proposal, in opposition to rights-based theorists like Dworkin (1977) and Mackie (1978), did not take rights to be morally basic. The proposal can be applied if we can presume that we have some independent grasp on the idea of the interests of members of a certain group, whether this is a preference-based conception of utility or a more Aristotelian conception of what the members of a group need in order to fulfil their nature as members of that group – i.e. to flourish. Then the rules that should govern the members of a group can be determined according to the principle that these should be the rules that enable the members of the group to flourish as such. These rules determine duties for the members of the group, and these duties determine specific rights for the members of the group.

This sort of discussion of rights, although it belongs squarely in moral philosophy, was brought into the remit of political philosophy as it developed in the second half of the century. Political philosophy’s concern with the ethical foundation of institutional systems, tax systems, judicial systems and so on, saw the role of rights – whether property rights (Nozick, 1974, Steiner, 1994) – or rights based on freedom – to be central in this discussion.

Even more central was ‘contractualism’ – the idea that actual or hypothetical social contracts should determine the right political systems and ground the individual’s moral commitment to such systems. Thomas Scanlon (1982) applied a ‘contractualist’ approach to individual morality by arguing that an act “is wrong if its performance under the circumstances would be disallowed by any system of rules for the general regulation of behaviour which no one could reasonably reject as a basis for informed, unforced general agreement” (1982, 110). And this kind of approach, often developed in a Kantian spirit as an exploration of the constraints that a certain conception of rationality imposed on moral systems, was a major force in political philosophy in the twentieth century.
Two of the most influential figures in political philosophy, John Rawls (1958) and Jürgen Habermas (1984), were neo-Kantians in this sense and tried to develop foundations for determining just institutional systems by describing what would count as a rational or reasonable way to agree on such things. Rawls argued that it would be reasonable to accept a system that we would all agree on were we deciding the matter from behind a veil of ignorance where we had no idea where we ourselves would figure in the system. And Habermas argued that it would be reasonable to accept a system that was agreed upon in a properly inclusive debate, where everyone had to respect everyone else’s perspective.

But at this point the narrative moves across to a discussion of Political Philosophy in the Twentieth Century, which is taken on by Chandran Kukathas in chapter ??.

References


Geach, P. 1958, “Imperative and deontic logic”, *Analysis* 18, 49-56.


van Inwagen, P. 1975, “The incompatibility of free will and determinism”, *Philosophical Studies* 27, 185-199.


