How cold the vacancy
When the phantoms are gone and the shaken realist
First sees reality. The mortal no
Has its emptiness and tragic expirations.
The tragedy, however, may have begun,
Again, in the imagination’s new beginning,
In the yes of the realist spoken because he must
Say yes, spoken because under every no
Lay a passion for yes that had never been broken.

WALLACE STEVENS, “Esthétique du Mal”

Quand on n’a pas de caractère, il faut bien se
donner une méthode.

ALBERT CAMUS, La Chute

CHAPTER 1

Socrates’ Question

It is not a trivial question, Socrates said: what we are talking about is how one should live. Or so Plato reports him, in one of the first books written about this subject. Plato thought that philosophy could answer the question. Like Socrates, he hoped that one could direct one’s life, if necessary redirect it, through an understanding that was distinctively philosophical — that is to say, general and abstract, rationally reflective, and concerned with what can be known through different kinds of inquiry.

The aims of moral philosophy, and any hopes it may have of being worth serious attention, are bound up with the fate of Socrates’ question, even if it is not true that philosophy, itself, can reasonably hope to answer it. With regard to that hope, there are two things to be mentioned here at the outset. One is particularly to be remembered by the writer — how large a claim he is making if he says that a particular kind of abstract, argumentative writing should be worth serious attention when these large questions are at issue. There are other books that bear on the question — almost all books, come to that, which are any good and which are concerned with human life at all. That is a point for the philosophical writer even if he does not think his relation to Socrates’ question lies in trying to answer it.

The other initial point is one for the reader. It would be a serious thing if philosophy could answer the question. How could it be that a subject, something studied in universities (but not only
there), something for which there is a large technical literature, could deliver what one might recognize as an answer to the basic questions of life? It is hard to see how this could be so, unless, as Socrates believed, the answer were one that the reader would recognize as one he might have given himself. But how could this be? And how would this be related to the existence of the subject? For Socrates, there was no such subject; he just talked with his friends in a plain way, and the writers he referred to (at least with any respect) were the poets. But within one generation Plato had linked the study of moral philosophy to difficult mathematical disciplines, and after two generations there were treatises on the subject—in particular, Aristotle's *Ethics*, still one of the most illuminating.²

Some philosophers would like to be able to go back now to Socrates' position and to start again, reflectively questioning common sense and our moral or ethical concerns, without the weight of texts and a tradition of philosophical study. There is something to be said for this, and in this book I shall try to follow it to the extent of pursuing an inquiry and hoping to involve the reader in it. At another level, however, it is baseless to suppose that one can or should try to get away from the practices of the subject. What makes an inquiry a philosophical one is reflective generality and a style of argument that claims to be rationally persuasive. It would be silly to forget that many acute and reflective people have already labored at formulating and discussing these questions. Moral philosophy has the problems it has because of its history and its present practices. Moreover, it is important that there is a tradition of activity, some of it technical, in other parts of philosophy, such as logic, the theory of meaning, and the philosophy of mind. While few of them outside mathematical logic provide "results," there is certainly a lot to be known about the state of the subject, and some of it bears significantly on moral philosophy.

There is another reason for not forgetting that we exist now and not in Socrates' condition. For him and for Plato it was a special feature of philosophy that it was reflective and stood back from ordinary practice and argument to define and criticize the attitudes involved in them. But modern life is so pervasively reflective, and a high degree of self-consciousness is so basic to its institutions, that these qualities cannot be what mainly distinguishes philosophy from other activities—from law, for instance, which is increasingly conscious of itself as a social creation; or medicine, forced to understand itself as at once care, business, and applied science; to say nothing of fiction, which even in its more popular forms needs to be conscious of its fictionality. Philosophy in the modern world cannot make any special claim to reflectiveness, though it may be able to make a special use of it.

This book will try to give some idea of the most important developments in moral philosophy, but it will proceed by way of an inquiry into its problems, in those directions that seem to me most interesting. I hope that the accounts of other people's work will be accurate, but they will assuredly be selective. It is not merely that my account of the subject will be different from one given by someone else (that must presumably be so if the book is worth reading at all), nor is it a question of how representative it will be, but rather that I shall not be concerned all the time to say how representative it is. There is one respect, at least, in which this book is not representative of ways in which the subject is for the most part now conducted, at least in the English-speaking world. It is more skeptical than much of that philosophy about what the powers of philosophy are, and it is also more skeptical about morality.

What the aims of moral philosophy should be depends on its own results. Because its inquiries are indeed reflective and general, and concerned with what can be known, they must try to give an account of what would have to go into answering Socrates' question: what part might be played by knowledge of the sciences; how far purely rational inquiry can take us; how far the answer to the question might be expected to be different if it is asked in one society rather than another; how much, at the end of all that, must be left to personal decision. Philosophical reflection thus has to consider what is involved in answering this, or any other less general, practical question, and to ask what powers of the mind and what forms of knowledge might be called upon by it. One thing that has to be considered in this process is the place of philosophy itself.

There might seem to be a circle in this: philosophy, in asking
how Socrates’ question might be answered, determines its own place in answering it. It is not a circle but a progression. Philosophy starts from questions that, on any view of it, it can and should ask, about the chances we have of finding out how best to live; in the course of that, it comes to see how much it itself may help, with discursive methods of analysis and argument, critical discontent, and an imaginative comparison of possibilities, which are what it most characteristically tries to add to our ordinary resources of historical and personal knowledge.

Socrates’ question is the best place for moral philosophy to start. It is better than “what is our duty?” or “how may we be good?” or even “how can we be happy?” Each of these questions takes too much for granted, although not everyone will agree about what that is. In the case of the last question, some people, such as those who want to start with the first question, will think that it starts in the wrong place, by ignoring the distinctive issues of morality; others may simply find it rather optimistic. Socrates’ question is neutral on those issues, and on many others. It would be wrong, however, to think that it takes nothing for granted. The first thing we should do is to ask what is involved in Socrates’ question, and how much we are presupposing if we assume that it can be usefully asked at all.

“How should one live?” — the generality of one already stakes a claim. The Greek language does not even give us one: the formula is impersonal. The implication is that something relevant or useful can be said to anyone, in general, and this implies that something general can be said, something that embraces or shapes the individual ambitions each person may bring to the question “how should I live?” (A larger implication can easily be found in this generality: that the question naturally leads us out of the concerns of the ego altogether. We shall come back to this later.) This is one way in which Socrates’ question goes beyond the everyday “what shall I do?” Another is that it is not immediate; it is not about what I should do now, or next. It is about a manner of life. The Greeks themselves were much impressed by the idea that such a question must, consequently, be about a whole life and that a good way of living had to issue in what, at its end, would be seen to have been a good life. Impressed by the power of fortune to wreck what looked like the best-shaped life, some of them, Socrates one of the first, sought a rational design of life which would reduce the power of fortune and would be to the greatest possible extent luck-free. This has been, in different forms, an aim of later thought as well. The idea that one must think, at this very general level, about a whole life may seem less compelling to some of us than it did to Socrates. But his question still does press a demand for reflection on one’s life as a whole, from every aspect and all the way down, even if we do not place as much weight as the Greeks did on how it may end.

The impersonal Greek phrase translated as one should is not only silent about the person whose life is in question. It is also entirely noncommittal, and very fruitfully so, about the kinds of consideration to be applied to the question. “How should I live?” does not mean “what life morally ought I to live?”; this is why Socrates’ question is a starting point different from those other questions I mentioned, about duty or about a life in which one would be good. It may be the same as a question about the good life, a life worth living, but that notion in itself does not bring in any distinctively moral claims. It may turn out, as Socrates believed and most of us still hope, that a good life is also the life of a good person (must be is what Socrates believed; can be is what most of us hope). But, if so, that will come out later. Should is simply should and, in itself, is no different in this very general question from what it is in any casual question, “what should I do now?”

Some philosophers have supposed that we cannot start from this general or indeterminate kind of practical question, because questions such as “what should I do?”, “what is the best way for me to live?”, and so on, are ambiguous and sustain both a moral and a nonmoral sense. On this view, the first thing one would have to do with the question is to decide which of these two different kinds of thing it meant, and until then one could not even start to answer it. That is a mistake. The analysis of meanings does not require “moral” and “nonmoral” as categories of meaning. Of course, if someone says of another “he is a good man,” we can ask whether the speaker means that he is morally good, as contrasted, for in-
stance, with meaning that he is a good man to take on a military sortie—but the fact that one can give these various interpretations no more yields a moral sense of "good" or of "good man" than it does a military sense (or a football sense, etc.).

One can of course ask, on a given occasion, "what should I do from an ethical point of view?" or "what should I do from a self-interested point of view?" These ask for the results of subdeliberations, and invite one to review a particular type of consideration among those that bear on the question and to think what the considerations of that type, taken by themselves, support. In the same way, I can ask what I should do taking only economic or political or family considerations into account. At the end of all that, there is the question "what should I do, all things considered?" There is only one kind of question to be asked about what to do, of which Socrates' is a very general example, and moral considerations are one kind of consideration that bear on answering it.4

Here and earlier I have mentioned "moral" considerations, using that word in a general way, which corresponds to what is, irremovably, one name for the subject: moral philosophy. But there is another name for the subject, "ethics," and corresponding to that is the notion of an ethical consideration. By origin, the difference between the two terms is that between Latin and Greek, each relating to a word meaning disposition or custom. One difference is that the Latin term from which "moral" comes emphasizes rather more the sense of social expectation, while the Greek favors that of individual character. But the word "morality" has by now taken on a more distinctive content, and I am going to suggest that morality should be understood as a particular development of the ethical, one that has a special significance in modern Western culture. It peculiarly emphasizes certain ethical notions rather than others, developing in particular a special notion of obligation, and it has some peculiar presuppositions. In view of these features it is also, I believe, something we should treat with a special skepticism. From now on, therefore, I shall for the most part use "ethical" as the broad term to stand for what this subject is certainly about, and "moral" and "morality" for the narrower system, the peculiarities of which will concern us later on.

I shall not try to define what exactly counts as an ethical consideration, but I shall say something about what goes into the notion of the ethical. It does no harm that the notion is vague. It is in fact morality, the special system, that demands a sharp boundary for itself (in demanding "moral" and "nonmoral" senses for words, for instance). This is a function of its special presuppositions. Without them, we can admit that there is a range of considerations that falls under the notion of the ethical, and we can also see why the range is not clearly delimited.

One thing that falls within its range is the notion of an obligation. A rather varied set of considerations is ordinarily counted as obligations, and I shall take up later (in Chapter 10) the question of why this should be so. One familiar kind is the obligation that one can put oneself under, in particular by making a promise. There is also the idea of a duty. The most familiar use of that word nowadays may be in narrow institutional connections, as when there is a list or roster of duties. Going beyond that, duties have characteristically been connected with a role, position, or relationship, such as those that follow from one's "station," as Bradley called it in the title of a famous essay.5 In a case such as the duties of a job, the job may have been acquired voluntarily, but in general duties, and most obligations other than those of promises, are not acquired voluntarily.

In the thought of Kant and of others influenced by him, all genuinely moral considerations rest, ultimately and at a deep level, in the agent's will. I cannot simply be required by my position in a social structure—by the fact that I am a particular person's child, for instance—to act in a certain way, if that was required is to be of the moral kind, and does not simply reflect a psychological compulsion or social and legal sanctions. To act morally is to act autonomously, not as the result of social pressure. This mirrors some of the characteristic concerns of the subsystem morality. As against that, it has been in every society a recognizable ethical thought, and remains so in ours, that one can be under a requirement of this kind simply because of who one is and of one's social situation. It may be a kind of consideration that some people in Western societies now would not want to accept, but it has been accepted by almost everyone in the past, and there is no necessity in the demand that
every requirement of this kind must, under rational scrutiny, be either abandoned or converted into a voluntary commitment. Such a demand is, like other distinctive features of morality, closely related to processes of modernization: it represents an understanding in ethical terms of the process that in the world of legal relations Maine called the change from status to contract. It corresponds also to a changing conception of the self that enters into ethical relations.\(^6\)

Obligation and duty look backwards, or at least sideways. The acts they require, supposing that one is deliberating about what to do, lie in the future, but the reasons for those acts lie in the fact that I have already promised, the job I have undertaken, the position I am already in. Another kind of ethical consideration looks forward, to the outcomes of the acts open to me. "It will be for the best" may be taken as the general form of this kind of consideration. In one way of taking this, specially important in philosophical theory, the best is measured by the degree to which people get what they want, are made happy, or some similar consideration. This is the area of welfarism or utilitarianism (I shall discuss such theories in Chapters 5 and 6). But that is only one version. G. E. Moore also thought that the forward-looking type of consideration was fundamental, but he allowed things other than satisfaction—such as friendship and the awareness of beauty—to count among the good consequences. It was because of this that his theory was so attractive to the Bloomsbury group: it managed to reject at once the stuffiness of duty and the vulgarity of utilitarianism.

There is another kind of ethical consideration, which presents an action as being of some ethically relevant kind. There is a wide range of ethical characteristics of actions under which they may be chosen or rejected. A particular action may be refused because it would be theft or murder, for instance, or deceitful or dishonorable, or, less dramatically, because it would let someone down. These descriptions—and there are many of them—operate at different levels; thus an action can be dishonorable because it is deceitful.

Closely connected with these descriptions, under which actions may be chosen or rejected, are various virtues, a virtue being a disposition of character to choose or reject actions because they are of a certain ethically relevant kind. The word "virtue" has for the most part acquired comic or otherwise undesirable associations, and few now use it except philosophers, but there is no other word that serves as well, and it has to be used in moral philosophy. One might hope that, with its proper meaning reestablished, it will come back into respectable use. In that proper use, meaning an ethically admirable disposition of character, it covers a broad class of characteristics, and, as so often in these subjects, the boundary of that class is not sharp and does not need to be made sharp. Some desirable personal characteristics certainly do not count as virtues, such as being sexually attractive. That can be a matter of character (some people have a sexually attractive character), but it does not have to be and it does not rate as a virtue, any more than having perfect pitch does. Again, virtues are always more than mere skills, since they involve characteristic patterns of desire and motivation. One can be a good pianist and have no desire to play, but if one is generous or fair-minded, those qualities themselves help to determine, in the right contexts, what one will want to do.

This is not to say that virtues can never be misused. One kind of virtue that can evidently be misused is the so-called executive virtues, which do not so much involve objectives of their own as assist in realizing other objectives—courage, for instance, or self-control. These are nevertheless virtues, being traits of character, and they are not related to pursuing other objectives as the mere possession of a skill is. According to Socrates, the virtues cannot be misused, and indeed he held something even stronger, that it is impossible for people, because they have a certain virtue, to act worse than if they did not have it. This led him, consistently, to believe that there is basically only one virtue, the power of right judgment. We need not follow him in that. More important, we should not follow him in what motivates those ideas, which is the search for something in an individual's life that can be unqualifiedly good, good under all possible circumstances. That search has its modern expressions as well, and we shall encounter one of them in the special preoccupations of morality.

The notion of a virtue is a traditional one in moral philosophy,
but it fell out of discussion for some time. In recent work, several writers have rightly emphasised its importance. If one has a certain virtue, then that affects how one deliberates. We need to be clear, however, about the ways in which it can affect the deliberation. An important point is that the virtue-term itself usually does not occur in the content of the deliberation. Someone who has a particular virtue does actions because they fall under certain descriptions and avoids others because they fall under other descriptions. That person is described in terms of the virtue, and so are his or her actions: thus he or she is a just or courageous person who does just or courageous things. But—and this is the point—it is rarely the case that the description that applies to the agent and to the action is the same as that in terms of which the agent chooses the action. “Just” is indeed such a case, one of the few, and a just or fair person is one who chooses actions because they are just and rejects others because they are unjust or unfair. But a courageous person does not typically choose acts as being courageous, and it is a notorious truth that a modest person does not act under the title of modesty. The benevolent or kindhearted person does benevolent things, but does them under other descriptions, such as “she needs it,” “it will cheer him up,” “it will stop the pain.” The description of the virtue is not itself the description that appears in the consideration. Moreover, there is typically no one ethical concept that characterizes the deliberations of a person who has a particular virtue. Rather, if an agent has a particular virtue, then certain ranges of fact become ethical considerations for that agent because he or she has that virtue. The road from the ethical considerations that weigh with a virtuous person to the description of the virtue itself is a tortuous one, and it is both defined and pitted by the impact of self-consciousness.

That same impact, in fact, may have contributed to making the virtues unpopular as an ethical conception. Their discussion used to make much of the cultivation of the virtues. In third-personal form, that exercise, if not under that title, is very familiar: it forms a good part of socialization or moral education or, come to that, education. As a first-personal exercise, however, the cultivation of the virtues has something suspect about it, of priggishness or self-deception. It is not simply that to think in this way is to think about oneself rather than about the world and other people. Some ethical thought, particularly if it is self-critical, will of course do that. More than one writer has recently stressed the importance of our capacity to have second-order desires—desires to have certain desires—and its significance for ethical reflection and the practical consciousness. Deliberation toward satisfying those second-order desires must be in a special degree directed toward the self. The trouble with cultivating the virtues, if it is seen as a first-personal and deliberative exercise, is rather that your thought is not self-directed enough. Thinking about your possible states in terms of the virtues is not so much to think about your actions, and it is not distinctively to think about the terms in which you could or should think about your actions: it is rather to think about the way in which others might describe or comment on the way in which you think about your actions, and if that represents the essential content of your deliberations, it really does seem a misdirection of the ethical attention. The lesson of all this, however, is not that the virtues are not an important ethical concept. It is rather that the importance of an ethical concept need not lie in its being itself an element of first-personal deliberation. The deliberations of people who are generous or brave, and also the deliberations of people who are trying to be more generous or braver, are different from the deliberations of those who are not like that, but the difference does not mainly lie in their thinking about themselves in terms of generosity or courage.

These, then, are some kinds of ethical concepts and considerations. What sorts of considerations bear on action but are not ethical considerations? There is one very obvious candidate, the considerations of egoism, those that relate merely to the comfort, excitement, self-esteem, power, or other advantage of the agent. The contrast between these considerations and the ethical is a platitude, and is grounded in obviously reasonable ideas about what ethical practices are for, the role they play in human societies. Yet even here distinctions need to be made. One is only a verbal point. We are concerned with Socrates’ question “how should one live?” and egoism, in the unvarnished and baldly self-interested sense, is at any rate an intelligible answer to that, even though most of us may be...
disposed to reject it. It is possible to use the word "ethical" of any scheme for living that would provide an intelligible answer to Socrates' question. In that sense, even the baldest egoism would be an ethical option. I do not think we should follow that use. However vague it may initially be, we have a conception of the ethical that understandably relates to us and our actions the demands, needs, claims, desires, and, generally, the lives of other people, and it is helpful to preserve this conception in what we are prepared to call an ethical consideration.

Egoism can, however, take a step farther than it takes in its baldest form. There is a theory of how we should act which has been called, confusingly enough, ethical egoism. This claims that each person ought to pursue his or her own self-interest. This differs from bald egoism because it is a reflective position and takes a general view about people's interests. Whether we call it an ethical system, as it calls itself, does not really matter very much. The important question is how it contributes to the idea of an ethical consideration. At first sight it seems to make no contribution to that, since it says that each of us ought to act on nonethical considerations. If it simply says that, it merely seems dogmatic: if people in fact act on considerations other than self-interest, what shows that they are irrational to do so? What this view is more likely to do, in fact, is to leave open the role of ethical considerations, and to ask how a life that involves acting on those considerations is related to self-interest.

There is another view, which looks much the same as the last but is different. It also claims something general, saying that what ought to happen is that everyone pursue his or her own interest. This view is likely to have an unstable effect on the considerations that one takes into account in acting. It may introduce a consideration that is ethical in the ordinary sense. If I believe that what ought to happen is that people pursue their own interest, then one thing I may have reason to do is to promote that state of affairs, and this may involve my giving a helping hand to others in adopting that policy. Such a line of action may well conflict with my simply pursuing my own self-interest.

In fact, it is quite difficult to sustain the bare belief that what ought to happen is that people pursue their own interest. It is more natural to support this with another consideration, that it is for the best if everyone does that. This may take the form of saying such things as that attempts to be kind to others merely confuse the issue. Someone who argues like this (and believes it) actually accepts some other ethical considerations as well, for instance that it is a good thing if people get what they want, and believes in addition that the best way for as many people as possible to get as much as possible of what they want is that each person should pursue what he or she wants. This is, of course, what advocates of laissez-faire capitalism used to claim in the early nineteenth century. Some even claim it in the late twentieth century, in the face of the obvious fact that all economic systems depend on people in society having dispositions that extend beyond self-interest. Perhaps this contradiction helps to explain why some advocates of laissez-faire tend to give moralizing lectures, not only to people who are failing to pursue their own interest but to people who are.

We are contrasting ethical and egoistic considerations. But might not somebody want someone else's happiness? Of course. Then would not egoism, my pursuit of what I want, coincide with what is supposed to be an ethical type of consideration, the concern for someone else's happiness? Again yes, but it is not very interesting unless in some more general and systematic way egoistic and ethical considerations come together. That is a question we shall come to when we consider foundations in Chapter 3.

From all this it will be seen that the idea of the ethical, even though it is vague, has some content to it; it is not a purely formal notion. One illustration of this lies in a different kind of nonethical consideration, which might be called the counterethical. Counterethical motivations, a significant human phenomenon, come in various forms, shaped by their positive counterparts in the ethical. Malevolence, the most familiar motive of this kind, is often associated with the agent's pleasure, and that is usually believed to be its natural state; but there exists a pure and selfless malevolence as well, a malice transcending even the agent's need to be around to enjoy the harm that it wills. It differs from counterjustice, a whimsical delight in unfairness. That is heavily parasitic on its ethical
counterpart, in the sense that a careful determination of the just is needed first, to give it direction. With malevolence it is not quite like that. It is not that benevolence has to do its work before malevolence has anything to go on, but rather that each uses the same perceptions and moves from them in different directions. (This is why, as Nietzsche remarked, cruelty needs to share the sensibility of the sympathetic, while brutality needs not to.) Other counterethical motivations, again, are parasitic on the reputation or emotional self-image of the ethical rather than on its conclusions. This, as one would expect, can particularly involve the virtues. That an action would be cowardly is not often found by an agent to be a consideration in its favor, but it could be, and in a counterethical way, ministering to a masochism of shame.

I have touched on considerations of egoism and on considerations that go outside the self—of benevolence, for instance, or fairness. But there is a question that has proved very important to ethics of how far outside the self such considerations should range. Will it count as an ethical consideration if you consider the interests and needs only of your family or of your community or of the nation? Certainly such local loyalties have provided the fabric of people's lives and the forum, it seems right to say, of ethical life. However, there are some ethical demands that seem to be satisfied only by a universal concern, one that extends to all human beings and perhaps beyond the human race. This concern is particularly cultivated by the subsystem morality, to the extent that it is often thought that no concern is truly moral unless it is marked by this universality.

For morality, the ethical constituency is always the same: the universal constituency. An allegiance to a smaller group, the loyalties to family or country, would have to be justified from the outside inward, by an argument that explained how it was a good thing that people should have allegiances that were less than universal. (I shall consider in Chapters 5 and 6 the motives and perils of this kind of approach; and also different accounts that have been given of what the universal constituency is.) At a more everyday level (a less reflective one, the moral critic would say), the location of the ethical can move from one side to another of a given contrast.

Relative to my personal interest, the interests of the town or the nation can represent an ethical demand, but the interests of the town can count as self-interested if the demand comes from some larger identification. This is simply because the requirements of benevolence or fairness may always stake a claim against self-interest; we can represent a self-interest as much as I; and who we are depends on the extent of identification in a particular case, and on the boundaries of contrast.

I have mentioned several sorts of ethical consideration, and more than one kind of nonethical. Philosophy has traditionally shown a desire to reduce this diversity, on both sides of the divide. It has tended, first of all, to see all nonethical considerations as reducible to egoism, the narrowest form of self-interest. Indeed some philosophers have wanted to reduce that to one special kind of egoistic concern, the pursuit of pleasure. Kant, in particular, believed that every action not done from moral principle was done for the agent's pleasure. This needs to be distinguished from another idea, that all actions, including those done for ethical reasons, are equally motivated by the pursuit of pleasure. This theory, psychological hedonism, finds it hard to avoid being either obviously false or else trivially vacuous, as it becomes if it simply identifies with the agent's expected pleasure anything that the agent intentionally does. But in any case this theory makes no special contribution to a distinction between the ethical and the nonethical. If there were any true and interesting version of psychological hedonism, those actions that had nonethical motivations would not necessarily form any special class of pleasure-seeking activity. Kant's view, on the other hand, does contribute to the question, by holding that moral action is uniquely exempted from psychological hedonism; that view is certainly wrong. If we are not influenced by such a theory, we can accept the obvious truth that there are different sorts of nonethical motivation—and, moreover, that there is more than one kind of motivation acting against ethical considerations.

The desire to reduce all nonethical considerations to one type is less strong in philosophy now than it was when moral philosophy chiefly concentrated not so much on questions of what is the right
thing to do and what is the good life (the answers to such questions were thought to be obvious), but rather on how one was to be motivated to pursue those things, against the motivations of selfishness and pleasure. The desire to reduce all ethical considerations to one pattern is, on the other hand, as strong as ever, and various theories try to show that one or another type of ethical consideration is basic, with other types to be explained in terms of it. Some take as basic a notion of obligation or duty, and the fact that we count it as an ethical consideration, for instance, that a certain act will probably lead to the best consequences is explained in terms of our having one duty, among others, to bring about the best consequences. Theories of this kind are called "deontological." (This term is sometimes said to come from the ancient Greek word for duty. There is no ancient Greek word for duty: it comes from the Greek for what one must do.)

Contrasted with these are theories that take as primary the idea of producing the best possible state of affairs. Theories of this kind are often called "teleological." The most important example is that which identifies the goodness of outcomes in terms of people's happiness or their getting what they want or prefer. This, as I have already said, is called utilitarianism, though that term has also been used, for instance by Moore, for the more general notion of a teleological system. Some of these reductive theories merely tell us what is rational, or again most true to our ethical experience, to treat as the fundamental notion. Others are bolder and claim that these relations are to be discovered in the meanings of what we say. Thus Moore claimed that "right" simply meant "productive of the greatest good," Moore's philosophy is marked by an affectation of modest caution, which clogged his prose with qualifications but rarely restrained him from wild error, and this, as a claim about what the words mean, is simply untrue. More generally, if theories of this kind are offered descriptively, as accounts of what we actually take to be equivalent, they are all equally misguided. We use a variety of different ethical considerations, which are genuinely different from one another, and this is what one would expect to find, if only because we are heirs to a long and complex ethical tradition, with many different religious and other social strands.

As an enterprise that intends to be descriptive, like anthropology, the reductive undertaking is merely wrongheaded. It may have other aims, however. It may, at some deeper level, seek to give us a theory of the subject matter of ethics. But it is not clear why that aim, either, must encourage us to reduce our basic ethical considerations. If there is such a thing as the truth about the subject matter of ethics — the truth, we might say, about the ethical — why is there any expectation that it should be simple? In particular, why should it be conceptually simple, using only one or two ethical concepts, such as duty or good state of affairs, rather than many? Perhaps we need as many concepts to describe it as we find we need, and no fewer.

The point of trying to reduce our ethical concepts must be found in a different aim of ethical theory, which is not just to describe how we think about the ethical but to tell us how we should think about it. Later I shall argue that philosophy should not try to produce ethical theory, though this does not mean that philosophy cannot offer any critique of ethical beliefs and ideas. I shall claim that in ethics the reductive enterprise has no justification and should disappear. My point here, however, is merely to stress that the enterprise needs justifying. A good deal of moral philosophy engages unblinkingly in this activity, for no obvious reason except that it has been going on for a long time.

There is one motive for reductivism that does not operate simply on the ethical, or on the nonethical, but tends to reduce every consideration to one basic kind. This rests on an assumption about rationality, to the effect that two considerations cannot be rationally weighed against each other unless there is a common consideration in terms of which they can be compared. This assumption is at once very powerful and utterly baseless. Quite apart from the ethical, aesthetic considerations can be weighed against economic ones (for instance) without being an application of them, and without their both being an example of a third kind of consideration. Politicians know that political considerations are not all made out of the same material as considerations against which they are weighed; even different political considerations can be made out of different material. If one compares one job, holiday, or companion with another, judgment does not need a particular set of weights.

This is not merely a matter of intellectual error. If it were that, it
could not survive the fact that people's experience contradicts it, that they regularly arrive at conclusions they regard as rational, or at least as reasonable, without using one currency of comparison. The drive toward a rationalistic conception of rationality comes instead from social features of the modern world, which impose on personal deliberation and on the idea of practical reason itself a model drawn from a particular understanding of public rationality. This understanding requires in principle every decision to be based on grounds that can be discursively explained. The requirement is not in fact met, and it probably does little for the aim that authority should be genuinely answerable. But it is an influential ideal and, by a reversal of the order of causes, it can look as if it were the result of applying to the public world an independent ideal of rationality. As an ideal, we shall see more of it later.13

Let us go back to Socrates' question. It is a particularly ambitious example of a personal practical question. The most immediate and uncomplicated question of that sort, by contrast, is “what am I to do?” or “what shall I do?” The various ethical and nonethical considerations we have been discussing contribute to answering such a question. Its answer, the conclusion of the deliberation, is of the form “I shall do . . .” or “what I am going to do is . . .”—and that is an expression of intention, an intention I have formed as a result of my deliberation. When it comes to the moment of action, it may be that I shall fail to carry it out, but then that will have to be because I have forgotten it, or been prevented, or have changed my mind, or because (as I may come to see) I never really meant it—it was not the real conclusion of my deliberation, or it was not a real deliberation. When the time for action is immediate, there is less room for these alternatives, so it is paradoxical if I come out with an answer of this kind and immediately fail to do what I said I was immediately going to do.

The question “what should I do?” allows rather more space between thought and action. Here the appropriate conclusion is “I should do . . .” and there are several intelligible ways of adding here “. . . but I am not going to.” Should draws attention to the reasons I have for acting in one way rather than another. The usual function of “I should . . . but I am not going to” is to draw attention to some special class of reasons, such as ethical or prudential reasons, which are particularly good as reasons to declare to others—because they serve to justify my conduct, for instance by fitting it into someone’s plan of action—but which are not, as it turns out, the strongest reasons for me, now; the strongest reason is that I desire very much to do something else. Desiring to do something is of course a reason for doing it.14 (It can even be a reason that justifies my conduct to others, though there are some tasks of justification, those particularly connected with justice, which by itself it cannot do.) So, in this sort of case, what I think I have most reason to do, taking all things together, is the thing I very much desire to do, and if I should is taken to refer to what I have most reason to do, this is what I should do. There is a further and deeper question, whether I can intentionally and without compulsion fail to do even what I think I have most reason to do; this, from Aristotle’s name for the phenomenon, is known as the problem of akrasia.15

Socrates’ question, then, means “how has one most reason to live?” In saying earlier that the force of should in the question was just should, I meant that no prior advantage is built into the question for one kind of reason over another. In particular, there is no special consideration for respectable justifying reasons. If ethical reasons, for instance, emerge importantly in the answer, that will not be because they have simply been selected for by the question.

Nevertheless, there is a peculiar emphasis given to Socrates’ question in that it stands at a distance from any actual and particular occasion of considering what to do. It is a general question about what to do, because it asks how to live, and it is also in a sense a timeless question, since it invites me to think about my life from no particular point in it. These two facts make it a reflective question. That does not determine the answer, but it does affect it. Answering a practical question at a particular time, in a particular situation, I shall be particularly concerned with what I want then. Socrates’ question I ask at no particular time—or, rather, the time when I no doubt ask it has no particular relation to the question. So I am bound by the question itself to take a more general, indeed a
longer-term, perspective on life. This does not determine that I give the answers of long-term prudence. The answer to the question might be: the best way for me to live is to do at any given time what I most want to do at that time. But if I have a weakness for prudence, the nature of Socrates' question is likely to bring it out.

It is, moreover, anybody's question. This does not mean, of course, that when asked by some particular person, it is a question about anybody: it is a question about that particular person. But when the question is put before me in the Socratic way, to invite reflection, it is going to be part of the reflection, because it is part of the knowledge constituting it, that the question can be put to anybody. Once constituted in that way, it very naturally moves from the question, asked by anybody, "how should I live?" to the question "how should anybody live?" That seems to ask for the reasons we all share for living in one way rather than another. It seems to ask for the conditions of the good life—the right life, perhaps, for human beings as such.

How far must the very business of Socratic reflection carry the question in that direction, and with what effects on the answer? The timelessness of the reflection does not determine that the answer should favor prudence. Similarly, the fact that the reflective question can be asked by anyone should allow its answer to be egoistic. But if it is egoistic, it will be egoism of one kind rather than another—the general egoism, distinguished earlier, which says that all people should favor their own interests. This naturally invites the thought that, if so, then it must be a better human life that is lived in such a way. But if so (it is tempting to go on), then it must be better, in some impersonal or interpersonal sense, that people should live in such a way. Having been led to this impersonal standpoint, perhaps we can be required to look back from it, make our journey in the reverse direction, and even revise our starting point. For if it is not better from an impersonal standpoint that each person should live in an egoistic way, perhaps we have a reason for saying that each of us should not live in such a way, and we must, after all, give a nonegoistic answer to Socrates' question. If all that does indeed follow, then the mere asking of Socrates' reflective question will take us a very long way into the ethical world. But does it follow?

Practical thought is radically first-personal. It must ask and answer the question "what shall I do?" Yet under Socratic reflection we seem to be driven to generalize the I and even to adopt, from the force of reflection alone, an ethical perspective. In Chapter 4, we shall see whether reflection can take us that far. But even if it cannot, Socratic reflection certainly takes us somewhere. Reflection involves some commitment, it seems, and certainly philosophy is committed to reflection. So the very existence of this book must raise the double question of how far reflection commits us and why we should be committed to reflection. Socrates thought that his reflection was inescapable. What he meant was not that everyone would engage in it, for he knew that not everyone would; nor that anyone who started reflecting on his life would, even against his will, be forced by inner compulsion to continue. His thought was rather that the good life must have reflection as part of its goodness: the unexamined life, as he put it, is not worth living.

This requires a very special answer to his question, which, for him, gives the final justification for raising it in the first place. If my book is committed to raising the question, is it committed to answering it in such a way? Must any philosophical inquiry into the ethical and into the good life require the value of philosophy itself and of a reflective intellectual stance to be part of the answer?
A NOTE of urgency can sometimes be heard, even in otherwise unhurried writers, when they ask for a justification of morality. Unless the ethical life, or (more narrowly) morality, can be justified by philosophy, we shall be open to relativism, amoralism, and disorder. As they often put it: when an amoralist calls ethical considerations in doubt, and suggests that there is no reason to follow the requirements of morality, what can we say to him? But what can we say to him if there is a justification of morality? Well, we can put the justification before him. But why should he be expected to stay where we have put it? Why should he listen? The amoralist, or even his more theoretical associate the relativist, is represented in these writings as an alarming figure, a threat. Why should it make any difference to such a person whether there is a philosophical justification of the ethical life?

Once at least in the history of philosophy the amoralist has been concretely represented as an alarming figure, in the character of Callicles who appears in Plato's dialogue the Gorgias. Callicles, indeed, under the conventions of Platonic dialogue, engages in rational conversation and stays to be humbled by Socrates' argument (an argument so unconvincing, in fact, that Plato later had to write the Republic to improve on it). What is unnerving about him, however, is something that Plato displays and that is also the subject of the dialogue: he has a glistening contempt for philosophy itself, and it is only by condescension or to amuse himself that he stays to listen to its arguments at all.

That is not the point. The question is not whether he will be convinced, but whether he ought to be convinced.

But is it? The writers' note of urgency suggests something else, that what will happen could turn on the outcome of these arguments, that the justification of the ethical life could be a force. If we are to take this seriously, then it is a real question, who is supposed to be listening. Why are they supposed to be listening? What will the professor's justification do, when they break down the door, smash his spectacles, take him away?

In any case, even if there is something that the rest of us would count as a justification of morality or the ethical life, is it true that the amoralist, call him Callicles, ought to be convinced? Is it meant only that it would be a good thing if he were convinced? It would no doubt be a good thing for us, but that is hardly the point. Is it meant to be a good thing for him? Is he being imprudent, for instance, acting against his own best interests? Or is he irrational in a more abstract sense, contradicting himself or going against the rules of logic? And if he is, why must he worry about that? Robert Nozick has well raised the question of what force the charge of inconsistency has against the "immoral man":

Suppose that we show that some X he holds or accepts or does commits him to behaving morally. He now must give up at least one of the following: (a) behaving immorally, (b) maintaining X, (c) being consistent about this matter in this respect. The immoral man tells us, "To tell you the truth, if I had to make the choice, I would give up being consistent."

It is not obvious what a justification of the ethical life should try to do, or why we should need such a thing. We should ask a pretended justification three questions: To whom is it addressed? From where? Against what? Against what, first of all, since we must ask what is being proposed as an alternative to the ethical life. It is important that there are alternatives to it. "The amoralist" is the name of somebody. This helps to define these questions in relation to a recurrent philosophical concern, skepticism. Skepticism can touch every kind of thing people claim to know: that there is an
"external" world; that other people have experiences (that there are other people, one may also say); that scientific inquiry can yield knowledge; that ethical considerations have force. Philosophical skepticism touches all these things, but in very different ways and with very different effects. In the case of the external world, the real question raised by skepticism, for any sane person, is not whether any of what we say about the world is true, or even whether we know any of it to be true, but how we know any of it to be true, and how much. There is no alternative within life to such beliefs: any alternative would have to be an alternative to life. In the case of "other minds," as that problem is often called, much the same is true, within the limits of sanity, but the problem shifts disquietingly toward how much? Certainly we know that other people have feelings, but how much do we know about those feelings? This is, in part, a philosophical question, one that has more practical effect than the mere question "how do I know?"

Ethical skepticism, in these respects, is at the opposite end of a line from skepticism about the external world. It is not, on the other hand, like skepticism about psychical research or psychoanalysis, where a real doubt is raised that might come eventually to be accepted, with the result that these activities would meet the same fate as phrenology: we would come to reject them altogether, finding that their claims to knowledge or even reasoned belief were baseless. It is not possible for ethical considerations to meet a collective rejection of that sort. For the individual, however, there does seem to be an alternative to accepting ethical considerations. It lies in a life that is not an ethical life.

Ethical skepticism of this sort differs so much from skepticism about the external world that it cannot be treated by the same methods. Moore famously disconcerted the skeptic about material objects by confronting him with one, Moore's hand (at any rate, it would have been a confrontation if such a skeptic had been there). There has been much discussion about the effect of Moore's gesture—as, for instance, whether it begged the question—but it undoubtedly has some effect, in reminding us that to take such a skeptic seriously might be to take him literally, and that there is some problem about what counts as doing that. There is no analogy here to the ethical. It may possibly be that if there are any ethical truths, some of them can be displayed as certain: given the choice, say, one should not surgically operate on a child without an anesthetic; but the production of such as example does not have the same disquieting effect on the ethical skeptic as the display of Moore's hand on the other kind. For one thing, one detached proposition known to be true about a material object will finish that first kind of skeptic: Moore's hand is an example of a material object, and, as one refutes none, so certainly one refutes possibly none. But the example of the child, or any other detached case of a striking sort, will count as an example of the ethical only to one who recognizes the ethical. The amoralist, Callicles himself if it took him that way, could help or spare a child. A limited benevolent or altruistic sentiment may move almost anyone to think that he should act in a certain way on a given occasion, but that fact does not present him with the ethical, as Moore's hand presented the skeptic with something material. The ethical involves more, a whole network of considerations, and the ethical skeptic could have a life that ignored such considerations altogether.

The traditional skeptic was basically a skeptic about knowledge, but an ethical skeptic is not necessarily the same as someone who doubts whether there is any ethical knowledge. In my sense, to be skeptical about ethics is to be skeptical about the force of ethical considerations; someone may grant them force, and so not be a skeptic, but still not think that they constitute knowledge because he does not think that the point lies in their being knowledge. (For the question whether there is ethical knowledge, see Chapter 8.) But, even when ethical skepticism is taken in this way, we should not assume that the skeptic must be someone who leads a life that goes against ethical considerations. Perhaps we should rather say that he leaves room for such a life. A skeptic, after all, is merely skeptical. As far as possible, he neither asserts nor denies, and the total skeptic, the Pyrrhonian of antiquity, was supposed neither to assert nor to deny anything. He could not bring it off, and it is doubtful that the ethical skeptic could bring that off—engage himself to use the ethical vocabulary, but with regard to every ethical question, suspend judgment. There are difficulties in the
very idea of doing that. It is hard, for instance, to use the vocabulary of promising and at the same time to sustain the position that there is nothing decisive to be said, for or against, on the question whether one ought to keep promises. Moreover, the skeptic has to act, and if he includes himself in the world of ethical discourse at all, then what he does must be taken as expressing thoughts he has within that world. If he speaks in terms of actions being ethically all right or not, and he cheerfully does a certain action, then we must take him to regard it as all right. So this is not an option for ethical skepticism. But there is another, which is to opt out of using ethical discourse altogether, except perhaps to deceive. While it is not an easy thing to do, the skeptic might be able to establish himself as one who is not at all concerned with ethical considerations. One can then see the force of the point that there are alternatives. He is not left with nothing to do.

The motivations the amoralist could be left with constitute one thing that the ethical claims might seek a justification against. Yet it is a mistake (as we shall see in the next chapter) to think that there is some objective presumption in favor of the nonethical life, that ethical skepticism is the natural state, and that the person we have been imagining is what we all would want to be if there were no justification for the ethical life and we had discovered that there was none. The moral philosopher in search of justifications sometimes pretends that this is so, overestimating in this respect the need for a justification just as he had overestimated its effect — its effect, at least, on the practicing skeptic.

This returns us to the question of “to whom?” When the philosopher raised the question of what we shall have to say to the skeptic or amoralist, he should rather have asked what we shall have to say about him. The justification he is looking for is in fact designed for the people who are largely within the ethical world, and the aim of the discourse is not to deal with someone who probably will not listen to it, but to reassure, strengthen, and give insight to those who will. This puts into a different perspective the idea we saw rather optimistically deployed in the case of the amoralist, that a justification of the ethical would be a force. Plato, who saw more deeply than any other philosopher into the questions raised by the possibility of a life outside the ethical, did not himself take it for granted that a justification of the ethical life would be a force. He thought that the power of the ethical was the power of reason, and that it had to be made into a force. He saw it as a problem of politics, and so it is. But he believed that the justification was intellectual and very difficult and, further, that everyone had some natural inclination to break out of the ethical order and destroy it. This inclination was a constant presence in most people, who lacked the capacity to master the justification and hence themselves. For Plato, the political problem of making the ethical into a force was the problem of making society embody the rational justification, and that problem could only have an authoritarian solution. If, by contrast, the justification is addressed to a community that is already an ethical one, then the politics of ethical discourse, including moral philosophy, are significantly different. The aim is not to control the enemies of the community or its shirkers but, by giving reason to people already disposed to hear it, to help in continually creating a community held together by that same disposition.

So far I have assumed for the most part that if we can engage in rational argument with someone, then we and that person are both within some ethical life (though not necessarily the same one): people outside any ethical life are unlikely to argue with us, and we have no great reason to trust them if they do. But that is not necessarily so. Leaving aside the desultory or, indeed, artificial motives that Callicles had for his conversation with Socrates, there is the important fact that people may be driven by a common need — at the limit, by a common fear of disaster — to negotiate understandings of limited cooperation or at least of nonaggression. There are inherent reasons why such agreements, without some external sanction, are bound to be unstable. In any case, they do not in themselves issue in any shared ethical understanding. This is enough to show that people can have a rational discussion without sharing an ethical system. Perhaps, for a limited purpose, they could rationally discuss without any of them having an ethical system. Yet for the most part this is not possible, because rational conversation between two parties, as an actual event, needs some-
thing to hold it together. This may, of course, be some particular relationship that does not extend more generally to the ethical, but if it is not that, and not the condescension of Callicles or the needs shared by those in a common emergency, then it must involve some minimal trace of an ethical consciousness.

This brings out once more the platitude that not all members of a community can live outside ethical life. But one person may be able to live outside it. This leads us to a first-personal form of the question whether it is possible to justify ethical considerations from the ground up. An agent who is asking Socrates’ question may wonder whether he could come to have reason for the ethical life, granted only some minimal structure of action, desire, or belief. This agent does not have to be someone who actually possesses only the minimum: he does not have to be outside the ethical world trying to see whether there is a way into it. He may as well, indeed better, be someone in it, who is considering what kinds of reasons he has for being there. (Again, how he might understand his reflection will itself be affected by its results.) Here we have no problem with the question “to whom?” The important question now is the last of our original three, “from what?” What is the minimum this person is assumed to have? If he is trying to justify the ethical life from the ground up, what is the ground?

In another well-worn image, where is there an Archimedean point? That question is not only worn, but profoundly discouraging for any inquiry it is taken to represent. In the case of some inquiries, we are so familiar with the discouragement that we find it hard to imagine what could count as success. If we were now set the task of finding some position outside all our knowledge and belief from which we could validate them, we might not understand the idea enough even to recognize it as a task. As we have seen, it is not immediately display a commitment to the ethical. That is why Socrates’ question is not already an ethical one, and also why the amoralist or skeptic sees the possibility of a rational life outside ethical considerations. Still, perhaps that is merely how it seems before one has reflected enough. The question itself did not use any specifically ethical terms, and that remains a fact. Nonetheless, it might turn out that when we properly think about it, we shall find that we are committed to an ethical life, merely because we are rational agents. Some philosophers believe that this is true. If they are right, then there is what I have called an Archimedean point: something to which even the amoralist or the skeptic is committed but which, properly thought through, will show us that he is irrational, or unreasonable, or at any rate mistaken.

There are two basic types of philosophical venture that fit this pattern. One of them works from the minimal and most abstract possible conception of rational agency. This will concern us in Chapter 4. The other, which we shall turn to immediately, assumes a richer and more determinate view of what rational agency is, taking it to be expressed in living a specifically human life. Both sets of ideas are rooted in past philosophies, the richer and more determinate conception in Aristotle, the more abstract in Kant. Neither of those philosophers, however, thinks that Socrates’ question can be taken simply as it stands, as a question, so to speak, waiting to be answered. Each of them redefines the search for an Archimedean point. They do so in different ways — but they have something important in common, which connects them with Socrates’ original questioning, as contrasted with other and less fruitful lines in the history of moral philosophy. Each of them yields an argument in practical reason. Neither aims in the first instance to prove the truth of some ethical proposition, which we are then asked to accept in virtue of our interest in believing the truth. Each of them rather commends certain action to us because of our interest in acting rationally or leading a satisfying human life. For both Aristotle and Kant, the justification of ethical propositions will come only from this, that they will be propositions accepted by one who acts rationally or leads such a life.
Socrates asks his question, in Plato’s Republic, in the course of a discussion with Thrasymachus, a fictional figure, it appears, created to embody some of the rougher assumptions of contemporary sophists. Thrasymachus concedes that one often does have a reason for being concerned with others’ interests as well as one’s own, but holds that this is only because one’s power is limited—typically, by the greater power of others. Naturally, according to Thrasymachus, human beings pursue power and pleasure. They may, rationally, have to curtail that pursuit because of other people’s power. They may also, irrationally, come to think that it is right or noble to respect others’ interests; but in that case they are being misled by conventions, social rules that inculcate these respectable but baseless assumptions. When they come to think like this, it is usually because, once more, someone else has greater power; their error is a deceit, and the conventions that deceive them are an instrument of coercion.

Thrasymachus says that the conventions that enjoin respect for others’ interests—“justice,” as it may be called— are an instrument of the strong to exploit the weak. This immediately raises the question, what makes these people strong? Thrasymachus speaks as if political or social power were not itself a matter of convention, and that is a view barely adequate to the school playground. His position is rapidly followed in the Republic by another, which takes this point. According to this, justice is the product of a convention adopted by a group of people to protect themselves. It is a contractual device of the weak to make themselves strong. This formulation is on the surface the opposite of the first one, and it is certainly more sophisticated, but the two have a good deal in common. By both views, justice is represented as an instrument for the satisfaction of selfish desires that exist naturally, independent of any ethical outlook. Both see justice as something one would not want to follow if one did not need to.

For Plato, this was a basic weakness. He thought that an account of the ethical life could answer Socrates’ question, and combat skepticism, only if it showed that it was rational for people to be just, whoever they were and whatever their circumstances. The second, contractual, account did no better in this respect than the original brutal view. If a man were powerful and intelligent and luckily enough placed, it would not be rational for him to conform to the conventional requirements of justice. The contractual theory was particularly weak in this respect because it was unstable with respect to a superior agent, one more intelligent and resourceful and persuasive than the rest. It was above all for this sort of agent that Plato thought skepticism had to be met, and justice and the ethical life shown to be rational.

In this respect, for Plato, the contractual theory failed. It failed, moreover, because of a certain structural feature: it represented as ethically basic a desirable or useful practice, the conventions of justice. But for Plato and for Socrates, what was first ethically desirable would have to be something that lay in the agent. If anything outside the soul, as they put it, is ethically primary—some rule, for instance, or institution—then we are left with the possibility that there could be a person whose deepest needs and the state of whose soul were such that it was not rational for him to act in accordance with that rule or institution and, so long as that was possible, the task of answering Socrates’ question in a way favorable to the ethical life would not be carried out. The demand to show to each person that justice was rational for that person meant that the answer had to be grounded first in an account of what sort of person it was rational to be.

It is sometimes said by modern critics that Platonic ethics—
and the same point arises with Aristotle's outlook as well—is egoistic, in a way that conflicts with the fundamental character of morality. The Greeks, it is suggested, had not arrived at a mature understanding of the moral consciousness. They had certainly not arrived at the distinctive preoccupations of the system morality, with its emphasis on a very special notion of obligation. (In this, as we shall see later, they were very fortunate.) But neither Plato nor Aristotle thought of the ethical life as a device that increased selfish satisfactions. Their outlook is formally egoistic, in the sense that they suppose that they have to show to each person that he has good reason to live ethically; and the reason has to appeal to that person in terms of something about himself, how and what he will be if he is a person with that sort of character. But their outlook is not egoistic in the sense that they try to show that the ethical life serves some set of individual satisfactions which is well defined before ethical considerations appear. Their aim is not, given an account of the self and its satisfactions, to show how the ethical life (luckily) fits them. It is to give an account of the self into which that life fits.

This is, already, a much more sophisticated objective than that of the crudest religious accounts, which represent ethical considerations as a set of laws or commands sanctioned by the promised punishments or rewards of God. This, crudest, level of religious morality is more egoistic. Even this kind of account, however, should not be dismissed on the ground that the egoistic motive it invokes could not possibly count, nor because we supposedly could not derive an ought from the fact of God's power but only from his goodness. There is nothing wrong with the general shape of this account: it explains why one would have good reason to live the kind of life that respected others' interests. It is rather that we know that it could not be true—could not be true, since if we understand anything about the world at all, we understand that it is not run like that. Indeed many, including many Christians, would now say we know that it is not run at all.

It is a natural thing to say that this religious account is crude, meaning not that it is crude because it is religious, but that it is a crude piece of religion. A less crude religious ethics will not add the religious element merely as an external sanction, but will give an account of human nature that provides equally for ethical objectives and for a relation to God. However, the criticism in terms of crude religion does raise a significant question. If religion is ultimately a matter of what the world is like, why should the world not be that crude? Why should religion be judged in terms of ethical understandings that are ranked as more or less sophisticated or mature in secular terms? The answer must be, presumably, that the original crude idea of God as an omnipotent law enforcer was itself gained through our (crude) ethical understanding. But then, if ethical understanding is going to develop, and if religion is going to understand its own development in relation to that, it seems inevitable that it must come to understand itself as a human construction; if it does, it must in the end collapse.

It is true that the development of the ethical consciousness means the collapse of religion, but not because a religious ethics, even a crude one, is logically debarred from being ethical. It is rather for a dialectical reason, that if the self-understanding of religion is not to be left behind by the ethical consciousness, it has to move in a direction that will destroy religion. The center of the matter does not lie in purely logical questions. In fact, the logical or structural questions about religious ethics, like many questions about God, are interesting only if you believe in God. If God exists, then arguments about him are arguments about the cosmos and of cosmic importance, but if he does not, they are not about anything. In that case, the important questions must be about human beings, and why, for instance, they ever believed that God existed. The issues about religious ethics are issues about the human impulses that expressed themselves in it, and they should be faced in those terms. For those who do not believe in a religious ethics, there is some evasion in continuing to argue about its structure: it distracts attention from the significant question of what such outlooks tell us about humanity. Nietzsche's saying, God is dead, can be taken to mean that we should now treat God as a dead person: we should allocate his legacies and try to write an honest biography of him.

Plato's aim, to return to that, was to give a picture of the self of such a kind that if people properly understood what they were, they
would see that a life of justice was a good not external to the self but, rather, an objective that it must be rational to pursue. For him, as for Aristotle, if it was rational to pursue a certain kind of life or to be a certain sort of person, then those things had to make for a satisfactory state called eudaimonia. That term is usually translated “happiness,” but what it refers to in the hands of these philosophers is not the same as modern conceptions of happiness. For one thing, it makes sense now to say that you are happy one day, unhappy another, but eudaimonia was a matter of the shape of one’s whole life. I shall use the expression well-being for such a state.

Socrates gave an account of it in terms of knowledge and the powers of discursive reason, and he could give this account because of the drastically dualistic terms in which he conceived of soul and body. Well-being was the desirable state of one’s soul—and that meant of oneself as a soul, since an indestructible and immaterial soul was what one really was. Such a conception underlay Socrates’ conception of our deepest interests and made it easier for him to believe that, in a famous phrase, the good man cannot be harmed, since the only thing that could touch him would be something that could touch the good state of his soul, and that was inviolable. It is a problem for this view that, in describing ethical motivations, it takes a very spiritual view of one’s own interests, but the subject matter of ethics requires it to give a less spiritual view of other people’s interests. If bodily hurt is no real harm, why does virtue require us so strongly not to hurt other people’s bodies?

There is another special consequence of Socrates’ picture, which relates to the hopes he had for the regenerative powers of philosophy. He, and to some extent Plato, believed that the discipline of philosophy could uniquely lead to well-being, through its power to develop the virtues. Rational philosophy was to provide the insight that led to well-being. This meant that philosophy either taught means to satisfy needs that were innate, or else it enabled us rationally to form a new conception of our needs. Aristotle’s outlook is less ambitious, and this is one reason, along with its much greater psychological and social elaboration, why Aristotle’s Ethics still serves as the paradigm of an approach that tries to base ethics on considerations of well-being and of a life worth living. For him, a human being is not an immaterial soul, but is essentially embodied and essentially lives a social life. Aristotle makes a basic distinction among the powers of reason, so that the intellectual faculty central to the ethical life, practical reason, is very different in its functions and objects from theoretical reason, which is what is deployed in philosophy and the sciences. He did indeed think that the cultivation of philosophy and sciences was the highest form of human activity, but he supposed that the exercise of practical reason in a personal and civic life was necessary to this, not only in the (Platonic) sense that such activities were necessary in society, but also in the sense that each individual needed such a life. The emphasis wobbles in Aristotle, though, between the civic life as a necessity the sage cannot escape and (what is certainly the more consistent and convincing consequence of his philosophical anthropology) as a necessity for each man if he is fully to express his powers.

Central to the life of practical reason are certain excellences of character or virtues, which are internalized dispositions of action, desire, and feeling. In some part, Aristotle’s account of the virtues, with regard to courage, for instance, or self-control, seems very recognizable; in other respects it belongs to another world. What matters for moral philosophy is whether the elements that are culturally more specific can be separated from the main structure. Some of them can be, and these include serious matters: an Aristotelian outlook is not committed to Aristotelian views on slavery or on the position of women. A center of doubt gathers, however, on the point that when Aristotle seems most removed from modern ethical perceptions, it is often because the admired agent is disquietingly concerned with himself. Aristotle does allow that the good man needs friends, and indeed that friendship is part of the good life; but he finds it necessary to argue for this in order to reconcile friendship with the ideal of self-sufficiency. Even his account of truthfulness has the oddly self-obsessed feature that the vices or faults contrasted with this virtue are not, as one would expect, untruthfulness (unreliability with regard to truth), but boasting and false modesty. We shall have to consider later whether the more self-centered aspects of Aristotle’s ethics stem from the structure itself.

I said that for Aristotle a virtue was an internalized disposition
of action, desire, and feeling. It is an intelligent disposition. It involves the agent's exercise of judgment, that same quality of practical reason, and so it is not simply a habit. It also involves favorable and unfavorable reactions to other people, their characters and actions. Aristotle's own views on this subject are bound up with one of the most celebrated and least useful parts of his system, the doctrine of the Mean, according to which every virtue of character lies between two correlative faults or vices (illustrated in the example of truth-telling), which consist respectively of the excess and the deficiency of something of which the virtue represents the right amount. The theory oscillates between an unhelpful analytical model (which Aristotle himself does not consistently follow) and a substantively depressing doctrine in favor of moderation. The doctrine of the Mean is better forgotten, but it does correctly imply that, since virtuous people are supposed to know what they are doing, they will see others' failings or vices as such and will see those who have them, or at least those people's actions, as variously bad or unpleasant or unhelpful or base.

Some of us are resistant to the idea that having a virtue or admirable disposition of character should also involve a disposition to assess others. The resistance has various roots, some of them very distant from any concern of Aristotle's. One is a conception of innocence, the image of a virtue that is entirely unselfconscious and lacking the contrast with self that is implied by judgment of others. Another root is skepticism, a suspicion that no one ever knows enough about anyone (including, in its more insidious versions, oneself) to make judgments. Still another is the fact that we accept, indeed regard as a platitude, an idea that Aristotle rejected, that someone can have one virtue while lacking others. For Aristotle, as for Socrates, practical reason required the dispositions of action and feeling to be harmonized; if any disposition was properly to count as a virtue, it had to be part of a rational structure that included all the virtues. This is quite different from our assumption that these kinds of disposition are enough like other psychological characteristics to explain how one person can, so to speak, do better in one area than another. This assumption, too, does something to inhibit reactions to other people.

Despite these considerations, there is still a connection between the ethical dispositions and reactions to others. The exact nature or depth of those reactions, and the degree of their self-confidence, will vary between individuals and in different cultural climates—but, as Aristotle claimed, an ethical disposition is not simply a personal pattern of behavior to which there may be contingently added a tendency to deplore or regret its absence in others. It is a kind of disposition that itself structures one's reactions to others. Because we do not believe in the unity of the virtues, we may accept the idea that it is simply a peculiarity of some people to lack certain ethical dispositions. It may even be that every ethical disposition can be seen in this way by someone, and none is so basic as to be exempt from being made, as one might put it, a subject of anecdote. But someone who sees every ethical disposition (or rather their subject matter, since he may not use those concepts himself) as a subject of anecdote surely lacks some basic ethical disposition. (Needless to say, it may be hard to find out whether someone does see them in that way.)

I have referred for the most part to the "reactions" to others that are involved in having ethical dispositions. It is a conveniently broad and unrevealing term, and there is much to be said about the range of attitudes, both positive and negative, that may fall under this heading. It is surprising how little of it has been said by moral philosophy, at least in the English-speaking tradition. By far the most important reason for this is the domination of morality, which is disposed to class all the relevant—that is to say, "moral"—reactions under headings such as judgment, assessment, and approval or disapproval. This is misleading in several ways. First, all these notions suggest a position of at least temporary superiority, the position of a judge, and this is so even if they occur within a moral theory that does not encourage superiority. Further, they suggest some binary judgment, as it might be of guilt or innocence. Moreover, they are supposedly directed only toward the voluntary: no one can properly attract moral judgment for what is not his fault. Because in this way it tries to cleave to an ultimate justice, morality does not merely provide a typology of reactions. It is not concerned simply with the question of what reactions are to be called moral.
The justice that is the aim of morality reaches further than the question of what your reactions should be called, to the issue of what reactions you may justly have, so that it comes to demand first a voice, then supremacy, and at last ubiquity. The “nonmoral” reactions such as dislike, or resentment or contempt, or such minor revelations of the ethical life as the sense that someone is creepy, are driven by a well-schooled moral conscience into a grumbling retreat, planning impersonation and revenge.

These various features of the moral judgment system support one another, and collectively they are modeled on the prerogatives of a Pelagian God. The strictness of the criteria for judgment responds to the supposed immensity of what is handed out, the finality of the only final justice there is. For the same reason, they collectively invite the skepticism I have mentioned. They face a problem of how people's character or dispositions could ever be the object of such a judgment. They are unlikely to be fully responsible for them, and it is even less likely that we can know to what extent they are responsible for them—even supposing we understand what we should know if we did know that.

Yet does morality require us to judge people's actions in isolation from their characters?

These are not Aristotelian worries. Aristotle did in fact think that human beings were in some absolute sense free, and that they brought forth their actions “like children.” He also thought that there were reactions of praise and blame directed to actions, and to people through those actions, which required the actions to be voluntary. But he would not have understood the suggestion that this was the limit to the reactions appropriate to others because of their ethical dispositions. In accepting, as we should, the idea that ethical dispositions are also dispositions to react, we must remember how much wider this range of reactions may be than is suggested by the conceptions of morality.

Aristotle should not have believed that in the most basic respects, at least, people were responsible for their characters. He gives an account of moral development in terms of habituation and internalization that leaves little room for practical reason to alter radically the objectives that a grown-up person has acquired.

Granted this conclusion, there is a problem about the way in which Aristotle presents his inquiry. Indeed, there is a problem about what he can take ethical philosophy to be. He presents it as a practical inquiry, one that is directed, in effect, to answering Socrates' question. He makes it seem as though you might review the whole of your life and consider whether it was aimed in the most worthwhile direction, but, on his own account, this cannot be a sensible picture. He shares with Plato the idea that, if virtue is part of human good, then it cannot be external to the ultimately desirable state of well-being; that state must be constituted in part by the virtuous life. But this is not a consideration that one could use to any radical effect in practical reasoning, as he seems to suggest. One becomes virtuous or fails to do so only through habituation. One should not study moral philosophy until middle age, Aristotle believes, for a reason that is itself an expression of the present difficulties—only by then is a person good at practical deliberation. But by then it will be a long time since one became, in relation to this deliberation, preemptively good or irrecoverably bad. (Only the powers of practical reason are in question here; it is consistent with everything Aristotle says that someone's life might be radically changed by other means, such as conversion.)

Some of Aristotle's reasonings might have an actual deliberative effect. He has an excellent argument about people who make the aim of their life political honor, that they tend to defeat themselves by making themselves dependent on those to whom they aim to be superior, and this discovery of Coriolanus' paradox, as it might be called, could surely serve some experienced person as a discovery or a diagnosis. But, in general, Aristotle cannot reasonably believe that his reflections on the virtuous life and its role in helping to constitute well-being could play a formative part in some general deliberation that a given person might conduct. In the light of this, the definition of ethical philosophy, and its aspirations, has to be revised. It no longer addresses its considerations to each person, so that each may answer the Socratic question. We come back to a point foreseen earlier, that the answer to the skeptic would be primarily for the benefit of the rest of us. Aristotle is in fact not interested in skepticism about the ethical life, and this is one of
many differences in urgency between his world and that of Socrates and Plato. He is concerned simply with men who have the wrong values or a bad character. But the point is the same, that the answer to Socrates’ question cannot be used by those who (from the perspective of the rest) most need it.

Still, this does not cast us to the opposite extreme, that the answer is simply meant to keep up the spirits of those within the system, give them more insight, and help them to bring up their children. The answer does that, but not only that. On Aristotle’s account a virtuous life would indeed conduce to the well-being of the man who has had a bad upbringing, even if he cannot see it. The fact that he is incurable, and cannot properly understand the diagnosis, does not mean that he is not ill. The answer Aristotle gives to Socrates’ question cannot be given to each person, as we have seen, but it is an answer for each person. Where exactly should we locate that thought? What exactly is being said about the bad man? We are not simply saying that we find him a dangerous nuisance (if we do), or that he is statistically unusual (if he is). We are saying that he lacks certain qualities characteristic of human beings which are necessary for creatures to live a life typical of human beings. But we have to say more, if we are to make the point essential to Aristotle’s philosophy and to any like it, that it is this man’s well-being and interests that are in question. We have to say that this man misconceives his interests and, indeed, that his doing so is a main symptom of what is wrong with him.

The notion that people may have “real interests” different from the interests they think they have is one that has generated a vast literature, and an almost equal amount of suspicion. The literature stems for the most part from the use of this notion made by Hegelian and, following Hegel, Marxist writers; the applications of the notion have been largely political; and in the light of those applications, the suspicions are largely well founded, since an appeal to people’s real interests is often deployed as a reason for coercing them contrary to their “apparent” (that is to say, perceived) interests. Some of these suspicions and criticisms, however, are wrongly directed at the notion of real interests itself. Even if a course of action were in someone’s real interests, the fact that it is not in his perceived interests does mean that, granted he cannot be persuaded, he will need to be coerced if his real interest is to be pursued. But in those circumstances, some further justification will be needed for our pursuing his real interests. It may be in Robinson’s real interests to stop drinking, but that does not instantly give anyone the right to stop him. (Who? — you? the doctor? the state?) The mere fact that real interests do not coincide with perceived interests already raises political and ethical issues.

The question of real interests in political thought raises further issues, in particular about class interests, which cannot be taken further here. The most general outlines of the problem, however, are clear. First, no controversial idea of real interests is involved if an agent merely lacks information which in the light of his other existing preferences and attitudes would alter his desires. He thinks that it is in his interest to drink this stuff because he believes it to be medicine prepared by his friendly pharmacist, but if it is actually cyanide, then he is certainly mistaken about his interests. The same applies to confusions of deliberative reasoning — though here there are pressing questions of what counts as a purely rational constraint on deliberative reasoning. Thus many philosophers think it irrational to prefer an earlier satisfaction to a later one just because it is earlier. (They admit that differences in certainty affect the issue in practice.) Others take it as obvious that the “proximity” of satisfactions, in Bentham’s phrase, is itself a dimension of practical reasoning. A conclusion on this point is certainly relevant to the question of what counts as a mistake in self-interested rationality.

The most significant questions about real interests arise when what is wrong with the agent goes beyond lack of information or mere rationality (whatever the boundaries of that may be) and affects the desires and motivations from which he deliberates; or, again, when what is wrong with the agent is that he will not believe something that he rationally should believe. A paradigm is the case of the despairing adolescent who attempts suicide (I mean that suicide is what he or she attempts, not that other thing, an attempted suicide). Susan, who has just attempted suicide, does not believe that things will look different in three months’ time, or if
she does believe that, she does not care — she does not want to be there for things to be better in three months’ time. If we believe that it will all be different in three months’ time, and we take steps to keep Susan alive, then it seems that we act in her real interest, an interest that, if we are right, she may well acknowledge in (say) six months’ time. That interest fails to be represented in her present motivations in a way that goes deeper than what has been discussed before. Susan’s lack of a desire to live, her disbelief in a better future, is itself part of the condition that will be cured in three months’ time. The inability to see what is in her interest is itself a symptom.

But we cannot simply say that a change is in someone’s real interest if, as a result of that change’s being made, she would acknowledge that it was in her interest. Perhaps, if you were to be brainwashed by a certain religious group, you would strongly identify your interests with those of the group. As a brainwashed believer, you might have much to say about an increase in enlightenment and the understanding you have now reached of your previous blindness — but that would not establish the value of brainwashing. Such difficulties arise with any psychological process that tends to generate belief in itself. One reaction to these difficulties is to give up and to regard the notion of real interests as incurably subjective or, perhaps, ideological. But a real problem remains, merely because there are some restrictions on what we can decently count as a certain person’s being better off as the result of a change, as opposed to things in general being better, or our being better off ourselves. “He would be better off dead” can be said for many dubious reasons: the most dubious is that we would be better off if he were dead.

If there is firmer footing to be found for the notion (and it seems that even the most skeptical treatment requires some further constraints), it will have to lie in the direction of excluding the self-validating changes, of the brainwashing type. A natural suggestion is the following. If an agent does not now acknowledge that a certain change would be in his interest and if, as a result of the change, he comes to acknowledge that it was in his interest, this will show that the change was really in his interest only on condition that the alteration in his outlook is explained in terms of some

general incapacity from which he suffered in his original state, and which has been removed or alleviated by the change. “General incapacity” is a vague phrase, but it carries two relevant ideas. One is that the agent’s alleged inability before the change to recognize his real interests is not simply tailor-made to the content of the recommended change, but has some more general implications, as the supposed inability to recognize the merits of the religious group did not. The second idea is that what is in question is indeed an incapacity. It is not simply that he does not acknowledge some things that he will acknowledge after the change, but that a capacity to acknowledge such things in cultural circumstances of that kind is to be expected in human beings, as part of their effective functioning. It is this last element, the normative conception of human functioning, that invited the terms “cure” and “symptom” in the description of the attempted suicide.

If we are going to bring in these notions eventually, why not do so earlier? Why not just say that a change is in someone’s real interest if the result of that change would be to bring him closer to normal human functioning? The answer is that not everything in someone’s interests is necessary to his human functioning, or is something that he needs. What he does need are the capacities, including the basic patterns of motivation, to pursue some of the things that are in his interests. If it is not to be purely ideological, the idea of real interests needs to be provided with a theory of error, a substantive account of how people may fail to recognize their real interests.12

Aristotle himself held a very strong theory of general teleology: each kind of thing had an ideal form of functioning, which fitted together with that of other things. He believed that all the excellences of character had to fit together into a harmonious self. Moreover, he was committed to thinking that the highest developments of human nature, which he identified with intellectual inquiry, would fit together with the more ordinary life of civic virtue, even though they represented the flowering of rather different powers, theoretical rather than practical reason. He was not very successful in showing this. Moreover, despite the rich teleological resources of his general account, he did not in fact do much to provide the
theory of error that the notion of real interests requires. He does
describe various kinds of bad people, and his descriptions are more
realistic than Plato's in the Republic, who (there, but not always)
gives in to the moralist's temptation to represent the bad person as a
compulsive addict, an unenviable wreck. Aristotle sees that some­
one can be in bad shape from the ethical point of view without
being at all like that—in particular, he may be able to use reason
effectively to pursue what he supposes is his advantage. Aristotle
explains that person's condition by saying that he was poorly
brought up, so that he acquired habits of pursuing the wrong kinds
of pleasure. But in Aristotle's teleological universe, every human
being (or at least every nondefective male who is not a natural slave)
has a kind of inner nisus toward a life of at least civic virtue, and
Aristotle does not say enough about how this is frustrated by poor
upbringing, to make it clear exactly how, after that upbringing, it is
still in this man’s real interest to be other than he is.

If Aristotle, with his strong assumptions about the nisus of
each natural kind of thing toward its perfection, cannot firmly
deliver this result, there is not much reason to think that we can.
Evolutionary biology, which gives us our best understanding of the
facts that Aristotle represented in terms of a metaphysical teleology,
cannot do better in trying to show that an ethical life is one of
well-being for each person. This is not because it delivers one an­
swer for all individuals, but one hostile to ethical life—for in­
stance, the answer that an entirely “hawkish” strategy would be
right for each and every individual. This is not so, since the out­
come would not constitute an “evolutionarily stable state,” as John
Maynard Smith has called it.13 The important point is that evolu­
tionary biology is not at all directly concerned with the well-being
of the individual, but with fitness, which is the likelihood of that
individual’s leaving offspring. The most that sociobiology might
do for ethics lies in a different direction, inasmuch as it might be
able to suggest that certain institutions or patterns of behavior are
not realistic options for human societies. That would be an impor­
tant achievement, but first sociobiology will have to be able to read
the historical record of human culture much better than it does
now.

If any science is going to yield conclusions that are for each
person, as I put it before, it will be some branch of psychology.
There are theories, particularly of a psychoanalytical kind, in which
hopes have been placed that they will support some ethical concep­
tion as a necessary part of human happiness. In some cases the
theories seem like this because they themselves involve what is
already ethical thought.14 They are none the worse for that, as
channels of individual help, and probably better, but this does
disqualify them from giving an independent account of well-being
and so providing a foundation for ethical life. Perhaps it is unrealis­
tic to suppose that there could be any psychological discipline
capable of doing this. It would be silly to try to determine a priori
and in a few pages whether there could be such a theory. It would
have to be at once independent of assumed ethical conceptions,
closely related to the complex aspects of human personality that are
involved in ethical life, determinate in its results, and—of course
—favorable to ethical considerations in some form. The last it
would “of course” have to be, not just for the boring reason that
only then would it count as providing foundations for ethical con­
siderations but because, if it failed to be favorable to ethical consid­
erations, it would have a different relation to practice altogether.

We need to live in society—and that is certainly an inner need, not
just a technological necessity—and if we are to live in society,
some ethical considerations or other must be embodied in the lives
of quite a lot of people. So a psychological theory which showed
that we could not really be happy in any adequate set of ethical
considerations would not tell us how to live: rather, it would predict
that we could not live happily.

Any adequate psychology of character will presumably include
the truth, in some scientifically presentable form, that many people
are horrible because they are unhappy, and conversely: where their
unhappiness is not something specially defined in ethical terms,
but is simply basic unhappiness—misery, rage, loneliness, despair.
That is a well-known and powerful fact; but it is only one in a range
of equally everyday facts. Some who are not horrible, and who try
hard to be generous and to accommodate others’ interests, are
miserable, and from their ethical state. They may be victims of a
suppressed self-assertion that might once have been acknowledged but now cannot be, still less overcome or redirected. There is also the figure, rarer perhaps than Callicles supposed, but real, who is horrible enough and not miserable at all but, by any ethological standard of the bright eye and the gleaming coat, dangerously flourishing. For those who want to ground the ethical life in psychological health, it is something of a problem that there can be such people at all. But it is a significant question, how far their existence, indeed the thought of their existence, is a cultural phenomenon. They seem sleeker and finer at a distance. Some Renaissance grandee fills such a role with more style than the tawdry fascist bosses, gangsters, or tycoons who seem, even as objects of fantasy, to be their chief contemporary instances. Perhaps we deceive ourselves about the past. Or perhaps it is an achievement of the modern world to have made it impossible to rear that type, because it has made evil, like other things, a collective enterprise, a process that makes it more powerful but less interesting. 15

Above all, in this table of naive perceptions, there is the matter of other goods. A certain man is horrible and rather miserable, but he is successful and has some pleasures, and if he were less horrible he would not be successful, and would be no less anxious, because he would be frustrated . . . Simply not to accept anything as valuable except the ethical dispositions—to turn, that is to say, the conception of psychological health in the direction of renouncing the other values—would be a reversion to Socratic asceticism and would need a reconstruction of the self to suit it. It would need also a utopian politics of renunciation by everyone; or else it would have to admit that virtue as purity of heart, while it was the only good, could be only a minority accomplishment, and this would need another politics in its turn, in order to construct the relation of that virtue to unregenerate society.

These problems take on a special significance, both for the individual and for the rest of society, when the “other goods” are of a creative and cultural kind. I have already said that it was a strain on Aristotle’s account of human nature to see such achievements as harmoniously of a piece with ordinary civic virtue. It is not of course a peculiarly modern thought, that it may not be possible to harmonise them; indeed Plato had taken a much more pessimistic view, and consequently had wished to banish the arts from the virtuous republic, or to domesticate them. But modern conceptions of the arts and the sciences and of the psychology of their creation can only make more intense, from these ethical perspectives, the problems of the wound and the bow, the unhappiness and the unloveliness that may be part of creative activity, which often has much to do with an imbalance, a hypertrophy of certain powers and sensibilities.

It is a problem also for any program that wants to connect the ethical life with psychological health through notions of integration, or reduction of conflict. These psychological aims in themselves cannot carry ethical weight unless they are already defined to do so; the best way of integrating some people would be to make them more ruthless. But apart from that, and also leaving aside those creative conflicts that raise doubts about how far conflict reduction may be psychologically desirable, there is a different sort of question, of how far and in what circumstances eliminating conflict may be ethically desirable. Conflict, in particular ethical conflict, may be the appropriate response to some kinds of situation. If these situations are to be eliminated, it will be a matter not only (perhaps not mainly) of reforming the psyche, but of changing society.

When one is considering the difficulties in psychology’s making a substantial contribution to the foundation of ethics, it is important to bear in mind how far we can go without it. We can go quite far. The formation of ethical dispositions is a natural process in human beings. This does not mean that it is spontaneous and needs no education or upbringing: in that sense, virtually nothing in human beings is “natural,” including the use of language—for while the capacity to learn a language is itself innate, and very probably specific, 16 no child will learn any language unless exposed to a particular language, which is itself, of course, a cultural product. Nor does it mean that the ethical life does not involve convention: it is natural to human beings to live by convention. There is no sense in which it is more natural, as Thrasymachus supposed, to live
outside ethical considerations. Moreover, we ourselves (most of us) are identified with some ethical considerations and have a conception of human well-being that gives a place to such considerations. We wish, consequently, to bring up children to share some of these ethical, as of other cultural, conceptions, and we see the process as good not just for us but for our children, both because it is part of our conception of their well-being and also because, even by more limited conceptions of happiness or contentment, we have little reason to believe that they will be happier if excluded from the ethical institutions of society. Even if we know that there are some people who are happier, by the minimal criteria, outside those institutions, we also know that they rarely become so by being educated as outlaws. As a result of all that, we have much reason for, and little reason against, bringing up children within the ethical world we inhabit, and if we succeed they themselves will see the world from the same perspective.

If we accept the displacement of Socrates' question implicit in what Aristotle admitted—that one cannot regenerate one's life by answering the question—we can, at one level, answer it with less than Aristotle offered. Displaced, it becomes a question about how we should live, and, at one level, we can give an ethical answer to that on the basis of the ethical life we have, even if we cannot claim, as Aristotle did, to have a teleological answer for each person, favoring an ethical life. At this level, the question will simply be whether society should be ethically reproduced, and to that question, merely from within society, we have an answer.

At this level. That argument presents only the choice between some ethical life and none. But ethical life is not a unitary given thing, and there are many different possibilities within it for education, social decision, even perhaps for personal regeneration. Within the kind of ethical life we find ourselves in, there are diversities, incoherences, and instruments of self-criticism. In our modern society there are more of all these, perhaps, than in any society that has ever existed, and this is a fact of the greatest importance, which changes the role of moral philosophy from anything imagined by an ancient writer. After I have considered, in the next chapter, a very different and more modern attempt to find foundations for the whole ethical enterprise, I shall go on to questions raised by these different possibilities for ethical life: some of them possibilities for different cultures, others for our own culture. We shall see what moral philosophy can do to help us understand them, and to ground self-criticism.

First, however, there is a question to be discussed about the extent of the distance we should acknowledge between Aristotle's conceptions and styles of ethical thought we might find acceptable now. In many substantial respects, as I have said, no modern discussion can share the outlook of an ancient writer. But how far does this extend to the logical shape of the whole enterprise? I said that, if the Socratic demand was to be met in its original form and the ethical life was to be justified to each person, then ethical value had to lie in some state of the self. The same would apply if, as I expressed Aristotle's aim, that justification could be given only for each person. But if we give up that objective as well, will ethical value still lie in states of the person? It is often thought to be a distinction between the ethics of the Greeks and modern conceptions that they approached ethical thought in this way and we do not. Moreover, that idea is associated with a deeper version of a criticism I mentioned before, that Greek ethical thought is incurably egoistic.

One kind of argument for this conclusion goes as follows. The person of Aristotelian virtue desires, quite often, to do various virtuous things. But anything motivated by desire is directed toward pleasure, and the pursuit of pleasure is egoistic. The only motivation opposed to this is the sense of obligation. Ethical motivation involves a contrast with the egoistic, so the ethical must be concerned with obligation, not with the desires that are involved in living a life of well-being, as Aristotle supposed. (This way of looking at things is a specially concentrated and crude version of the outlook of morality.) Almost all the assumptions of this argument are wrong. It is false, indeed incoherent, to suppose that every desire aims at pleasure, and it would be false even if the satisfaction of each desire issued in pleasure (which is not so either). Moreover, if it were true that every desire aimed at pleasure, one could not rely on the common-sense assumption that there is a contrast between
ethical motivations and pleasure seeking. Ethical motivations would then aim at certain sorts of pleasure. Some of them do indeed issue in pleasure, and Hume, in line with Greek thought on this point, agreeably thought that it was the mark of a virtuous person to take pleasure in doing generous or helpful actions.

One obvious reason why my desires do not all have as their object my pleasure is that some of my desires aim at states of affairs that do not involve me at all: I am not mentioned in a full specification of what would satisfy such a desire. There are self-transcending desires. They are not all altruistic or benevolent—they may be malicious or frivolous. Those who make provisions in their wills to mortify their relatives or to promote some absurd object do not usually believe that they will be there to enjoy the outcome; yet it is the outcome they want, not merely the pleasure of thinking about it now. For all these reasons, the line between self-concern and other-concern in no way corresponds to a line between desire and obligation. (Indeed, some moralists admit this in their own way by inventing a class of duties to oneself, self-regarding obligations. These serve a number of functions in that economy. One is to encourage long-term investment as against consumption; another is merely to launder the currency of desire.)

Even when we have got rid of these misconceptions about desire and pleasure, however, there may still seem to be something left to the charge of egoism. The ethical dispositions are dispositions to want certain things, to react in certain ways to other people and to their actions, to use such notions as that of obligation, to promote certain outcomes as being just, and so on. The agent will probably be a party to the relations involved, and of course it is the agent who asks and decides how he is going to act. None of these conceptions (including his wants) need damagingly involve the agent’s self in its content: none of this, in itself, involves any kind of egoism. But the Socratic question brings in another idea. It involves the agent’s thinking about these dispositions themselves and relating them to a life of well-being. Even if the dispositions are not themselves directed toward the self, it is still his own well-being that the agent in Socratic reflection will be considering. Egoism seems to be back again.

The answer to this problem lies in the vital fact that the Aristotelian account puts the substantive ethical dispositions into the content of the self. I am, at the time of mature reflection, what I have become, and my reflection, even if it is about my dispositions, must at the same time be expressive of them. I think about ethical and other goods from an ethical point of view that I have already acquired and that is part of what I am. In thinking about ethical and other goods, the agent thinks from a point of view that already places those goods, in general terms, in relation to one another and gives a special significance to ethical goods. Looked at from the outside, this point of view belongs to someone in whom the ethical dispositions he has acquired lie deeper than other wants and preferences.

The difference between the inside point of view, the view from one’s dispositions, and the outside view of those dispositions shows how it is that in the most obvious sense it is not true that all ethical value rests in the dispositions of the self, and yet, in another way, it is true. It is not true from the point of view constituted by the ethical dispositions—the internal perspective—that the only things of value are people’s dispositions; still less that only the agent’s dispositions have value. Other people’s welfare, the requirements of justice, and other things, have value. If we take up the other perspective, however, and look at people’s dispositions from the outside, we may ask the question “what has to exist in the world for that ethical point of view to exist?” The answer can only be, “people’s dispositions.” There is a sense in which they are the ultimate supports of ethical value. That has a practical as well as a metaphysical significance. The preservation of ethical value lies in the reproduction of ethical dispositions.

The outside point of view of his dispositions is available to the agent himself. But if he tries in his reflection to abstract himself totally from those dispositions, and to think about himself and the world as though he did not have them, then he should not be surprised if he cannot get an adequate picture of the value of anything, including his own dispositions. He cannot do so, precisely because those dispositions are part of the content of his actual self. Moreover, if he is to conduct any reflection in which he
stands back from his own dispositions, it is important whether there is anything in the view of things he takes from the outside that conflicts with the view of things he takes from the inside. For Aristotle, the virtuous agent would find no such conflict. He could come to understand that the dispositions that gave him his ethical view of the world were a correct or full development of human potentiality. This was so absolutely, in the sense at least (Aristotle no doubt meant more) that the best possible theory of humanity and its place in the world would yield this result. Also, this perfection could be displayed harmoniously, so that the development of these ethical capacities would fit with other forms of human excellence. Aristotle’s theory means that when the agent reflects, even from the outside, on all his needs and capacities, he will find no conflict with his ethical dispositions.

Here we meet again the many modern doubts that weaken this account. Our present understanding gives us no reason to expect that ethical dispositions can be fully harmonized with other cultural and personal aspirations that have as good a claim to represent human development. Even if we leave the door open to a psychology that might go some way in the Aristotelian direction, it is hard to believe that an account of human nature—if it is not already an ethical theory itself—will adequately determine one kind of ethical life as against others. Aristotle saw a certain kind of ethical, cultural, and indeed political life as a harmonious culmination of human potentialities, recoverable from an absolute understanding of nature. We have no reason to believe in that. Once we lose the belief, however, a potential gap opens between the agent’s perspective and the outside view. We understand—and, most important, the agent can come to understand—that the agent’s perspective is only one of many that are equally compatible with human nature, all open to various conflicts within themselves and with other cultural aims. With that gap opened, the claim I expressed by saying that agents’ dispositions are the “ultimate supports” of ethical value takes on a more skeptical tone. It no longer sounds enough.

I believe that the claim is true, and that in its general outline the description of the ethical self we have recovered from the ancients is correct. At the same time, we must admit that the Aristotelian assumptions which fitted together the agent’s perspective and the outside view have collapsed. No one has yet found a good way of doing without those assumptions. That is the state of affairs on which the argument of this book will turn, and I shall come back in various connections to the relations between the inside and the outside points of view. My next concern, however, is with a different attempt to start from the ground up, one that tries to find an Archimedean point without using Aristotelian assumptions.
CHAPTER 4
Foundations: Practical Reason

The project of the last chapter, which tried to ground ethical life in well-being, sought determinate conclusions about the shape of a whole life, from substantive beliefs about human nature. We saw that it needed very strong assumptions to hold it together, assumptions we cannot accept.

There is another project that also tries to start from the ground up but claims to deliver less, from less. Instead of giving an account of a fully developed life, it offers certain structural or formal features of ethical relations. Instead of relying on a specific teleology of human nature, it starts from a very abstract conception of rational agency. It still tries to give an answer to Socrates' question, though a minimal one. It gives the answer to each agent, merely because the agent can ask the question. Hence its answers are more abstract and less determinately human than those in the Aristotelian style. This type of argument yields, if anything, general and formal principles to regulate the shape of relations between rational agents. These are the concerns of Kant.

This may seem a surprising thing to say. Kant's name is associated with an approach to morality in which, it is often supposed, there can be no foundations for morality at all. He insisted that morality should be "autonomous," and that there could be no reason for being moral. A simple argument shows why, in the Kantian framework, this must be so. Any reason for being moral must be either a moral or a nonmoral reason. If it is moral, then it cannot really be a reason for being moral, since you would have to be already inside morality in order to accept it. A nonmoral reason, on the other hand, cannot be a reason for being moral; morality requires a purity of motive, a basically moral intentionality (which Kant took to be obligation), and that is destroyed by any nonmoral inducement. Hence there can be no reason for being moral, and morality presents itself as an unmediated demand, a categorical imperative.

It is specifically morality that Kant introduces, and we shall face wider questions about this conception of the ethical life when we come to that subject in Chapter 10. Kant's outlook indeed requires that there be no reason for morality, if that means a motivation or inducement for being moral, but it does not imply that morality has no foundations. Kant thought that we could come to understand why morality should rightly present itself to the rational agent as a categorical demand. It was because rational agency itself involved accepting such a demand, and this is why Kant described morality in terms of laws laid down by practical reason for itself.1

In his extraordinary book The Groundwork of the Metaphysic of Morals, the most significant work of moral philosophy after Aristotle, and one of the most puzzling, he tries to explain how this can be. I do not want to try to set out the argument, however, by directly expounding Kant. That would involve many special problems of its own. I shall treat his outlook as the destination rather than the route and shall develop in the first place an argument that will be simpler and more concrete than Kant's.2 The failure of that argument to give morality a foundation will help to show why the Kantian conception needs to be as metaphysically ambitious as it is. I do not believe that Kant's argument succeeds either, but one has to follow it a long way down to find out why not.

Is there anything that rational agents necessarily want? That is to say, is there anything they want (or would want if they thought hard enough about it) merely as part or precondition of being agents?

When they are going to act, people necessarily want, first of all, some outcome: they want the world to be one way rather than another. You can want an outcome without wanting to produce
that outcome — you might prefer that the outcome merely materialize. Indeed, there are some cases in which the outcome you want will count only if you do not directly produce it (you want her to fall in love with you). But, in direct contrast to that possibility, in many cases you essentially want not only the outcome, but to produce the outcome. To put it another way (a way that is complicated but still conceals some complications), the outcome you want itself includes the action that your present deliberation will issue in your doing. 4

We do not merely want the world to contain certain states of affairs (it is a deep error of consequentialism to believe that this is all we want). Among the things we basically want is to act in certain ways. But even when we basically want some state of affairs, and would be happy if it materialized, we know that we do not live in a magical world, where wanting an outcome can make it so. Knowing, therefore, that it will not come about unless we act to produce it, when we want an outcome we usually also want to produce it. (There is an direct analogy to this in the principle that, when we want the truth, we want to know the truth. 4) Moreover, we do not want it merely to turn out that we produced it; we want these thoughts of ours to produce it. The wants involved in our purposive activities thus turn out to be complex. At the very least, what we want is that the outcome should come about because we wanted it, because we believed certain things, and because we acted as we did on the basis of those wants and beliefs. 5 Similar considerations apply to keeping things that we want to keep.

This adds up, then, to the following: on various occasions we want certain outcomes; we usually want to produce those outcomes; we usually want to produce them in a way that expresses our want to produce them. Obviously enough, on those occasions we do not want to be frustrated, for instance by other people. Reflecting on all this, we can see that we have a general, dispositional, want not to be frustrated, in particular by other people. We have a general want, summarily put, for freedom. This is not to deny that sometimes we want to lose freedom, to be frustrated by others, even to be coerced — but then we do not want to be frustrated in obtaining that. 6

It is not enough, though, for this freedom merely that we should not be frustrated in doing whatever it is we want to do. We might be able to do everything we wanted, simply because we wanted too little. We might have unnaturally straitened or impoverished wants. This consideration shows that we have another general want, if an indeterminate one: we want (to put it vaguely) an adequate range of wants.

It does not follow from all this that we want our choices to be as little limited as possible, by anything or anyone. We do not want our freedom to be limitless. It may seem to follow, 7 but to accept it would be to leave out another vital condition of rational agency. Some things, clearly, are accessible to an agent at a given time and others are not. Moreover, what is accessible, and how easily, depends on features both inside and outside the agent. He chooses, makes up plans, and so on, in a world that has a certain practicable shape, in terms of where he is, what he is, and what he may become. The agent not only knows this is so (that is to say, he is sane), but he also knows, on reflection, that it is necessary if he is indeed going to be a rational agent. Moreover, he cannot coherently think that in an ideal world he would not need to be a rational agent. The fact that there are restrictions on what he can do is what requires him to be a rational agent, and it also makes it possible for him to be one; more than that, it is also the condition of his being some particular person, of living a life at all. We may think sometimes that we are dismally constrained to be rational agents, and that in a happier world it would not be necessary. But that is a fantasy (indeed it is the fantasy).

Similar conditions apply to the agent’s knowledge. Acting in a particular situation, he must want his plans not to go wrong through ignorance or error. But even in that particular case, he does not want to know everything, or that his action should have no unintended consequences. Not to know everything is, once more, a condition of having a life — some things are unknown, for instance, because they will form one’s future. If you cannot coherently want to know everything, then you also cannot coherently want never to be in error. They are not the same thing (omniscience is not the same as infallibility), but there are many connections
between them. For one thing, as Karl Popper has always emphasized, you must make errors, and recognize them, if you are going to extend such knowledge as you have.

These last considerations have concerned things a rational agent does not need to want, indeed needs not to want, as a condition of being such an agent. They assume him or her to be a finite, embodied, historically placed agent: the only kind of agent I take there to be, with the marginal or dubious exception of corporations and similar agencies, and (with the same exceptions) the only ones that could be the concern of ethics. (Even those who believe in God, though they take him to be an agent, should not take him to be the concern of ethics.) I suppose this is what most people would expect. But it has some important consequences, which will concern us later.

As rational agents, then, we want what I have summarily called freedom, though that does not mean limitless freedom. Does this commit us to thinking that our freedom is a good and that it is a good thing for us to be free? One path leading to this conclusion would be to say that when an agent wants various particular outcomes, he must think that those various outcomes are good. Then he would be bound to think that his freedom was a good thing, since it was involved in securing those outcomes. 8

Is it true that if we want something and purposively pursue it, then we think of our getting that thing as good? This is a traditional doctrine, advanced in Plato’s Meno and hallowed in a saying of scholastic philosophy, omne appetitum appetitur sub specie boni, everything pursued is pursued as something good. It seems to me not true. In any ordinary understanding of good, surely, an extra step is taken if you go from saying that you want something or have decided to pursue it to saying that it is good, or (more to the point) that it is good that you should have it. The idea of something’s being good imports an idea, however minimal or hazy, of a perspective in which it can be acknowledged by more than one agent as good. An agent who merely has a certain purpose may of course think that his purpose is good, but he does not have to. The most he would commit himself to merely by having a purpose would presumably be that it would be good for him if he succeeded in it, but must even this much be involved? Even this modest claim implies a perspective that goes somewhere beyond the agent’s immediate wants, to his longer-term interests or well-being. To value something, even relatively to your own interests, as you do in thinking that it would be better “for me,” is always to go beyond merely wanting something. I might indeed come to put all the value in my life into the satisfaction of one desire, but if I did, it would not simply be because I had only one desire. Merely to have one desire might well be to have no value in my life at all; to find all the value in one desire is to have just one desire that matters to me. 9

Even if we give up the traditional doctrine, however, so that I do not have to see everything I want as good, it might still be true that I should see my freedom as good. “Good for me,” I suggested, introduces some reference to my interests or well-being that goes beyond my immediate purposes, and my freedom is one of my fundamental interests. So perhaps I must regard my own freedom as a good. But if so, I must not be misled into thinking that my freedom constitutes a good, period. This would be so only if it were a good, period, that I should be a rational agent, and there is no reason why others should assent to that. In fact, it is not even clear that I have to assent to it. This begins to touch on some deeper questions about my conception of my own existence.

Everything said so far about the basic conditions and presuppositions of rational action seems to be correct. The argument that tries to provide a foundation for morality attempts to show that, merely because of those conditions, each agent is involved in a moral commitment. Each agent, according to this argument, must think as follows. Since I necessarily want my basic freedom, I must be opposed to courses of action that would remove it. Hence I cannot agree to any arrangement of things by which others would have the right to remove my basic freedom. So when I reflect on what arrangement of things I basically need, I see that I must claim a right to my basic freedom. In effect, I must lay it down as a rule for others that they respect my freedom. I claim this right solely because I am a rational agent with purposes. But if this fact alone is the basis of my claim, then a similar fact must equally be the basis of such a claim by others. If, as I suppose, I legitimately and appropri-
ately think that they should respect my freedom, then I must recognize that they legitimately and appropriately think that I should respect their freedom. In moving from my need for freedom to “they ought not to interferes with me,” I must equally move from their need to “I ought not to interfere with them.”

If this is correct, then each person’s basic needs and wants commit him to stepping into morality, a morality of rights and duties, and someone who rejects that step will be in a kind of pragmatic conflict with himself. Committed to being a rational agent, he will be trying to reject the commitments necessarily involved in that. But is the argument correct? Its very last step—that if in my case rational agency alone is the ground of a right to noninterference, then it must be so in the case of other people—is certainly sound. It rests on the weakest and least contestable version of a “principle of universalizability,” which is brought into play simply by because or in virtue of. If a particular consideration is really enough to establish a conclusion in my case, then it is enough to establish it in anyone’s case. That must be so if enough is indeed enough. If the conclusion that brings in morality does not follow, it must be because of an earlier step. Granted that the original claims are correct about a rational agent’s wants and needs, the argument must go wrong when I first assert my supposed right.

It is useful to consider what the agent might say in thinking out his claims. It could be put like this:

I have certain purposes.
I need freedom to pursue these or any other purposes.
So, I need freedom.
I prescribe: let others not interfere with my freedom.

Call the one who is thinking this, the agent A. Assume for the moment that we know what a “prescription” is, and call this prescription of A’s, Pa. Then A also thinks

Pa is reasonable,

where what this means is that Pa is reasonably related to his, A’s, being a rational agent. A can of course recognize that another agent, say B, can have thoughts just like his own. He knows, for instance, that

B prescribes: let A not interfere with my freedom,

and, calling B’s prescription Pb, the principle of universalizability will require A to agree that

Pb is reasonable.

It may look as if he has now accepted B’s prescription as reasonable in the sense of making some claim on himself. This is what the argument to morality requires. But A has not agreed to this. He has agreed only that Pb is reasonable in the same sense that Pa is, and what this means is only that Pb is reasonably related to B’s being a rational agent—that is to say, B is as rational in making his prescription as A is rational in making his. It does not mean that B would be rational in accepting Pa (or conversely) if in accepting it he would be committing himself not to interfere with A’s freedom.

The same point comes out in this: one could never get to the required result, the entry into the ethical world, just from the consideration of the should or ought of rational agency itself, the should of the practical question. The reasons that B has for doing something are not in themselves reasons for another’s doing anything. The should of practical reason has, like any other, a second and a third person, but these forms merely represent my perspective on your or his interests and rational calculations, the perspective of “if I were you.” Considering in those terms what B should do, I may well conclude that he should interfere with my freedom.

But can I “prescribe” this for him? What does it mean? Certainly I do not want him to interfere with my freedom. But does this, in itself, generate any prescription that leads to obligations or rights? The argument suggests that if I do not prescribe that others ought not to interfere with my freedom, I shall be logically required to admit that they may interfere with it—which I do not want to do. What the argument claims is that I must either give them the right to interfere with my freedom or withhold that right from them. The argument insists, in effect, that if I am to be consistent, I must
make a rule to the effect that others should not interfere with my freedom, and nothing less than this rule will do. But the rule, of course, just because it is a general rule, will equally require me not to interfere with their freedom.

But why must I prescribe any rule? If I am in the business of making rules, then clearly I will not make one enjoining others to interfere with my freedom, nor will I make one permitting them to do so. But there is another possibility: I do not regard myself as being in this business, and I make no rule either way. I do not have to be taken as giving permission. If there is a system of rules, then no doubt if the rules are silent on a certain matter (at least if the rules are otherwise wide enough in their scope), that fact can naturally be taken to mean permission. The law, like other sovereign agencies, can say something by remaining silent. But if there is no law, then silence is not meaningful, permissive, silence: it is simply silence. In another sense, of course, people "may" interfere with my freedom, but that means only that there is no law to stop, permit, or enjoin. Whether they "may" means they "can" depends on me and what I can do. As the egoist Max Stirner put it: "The tiger that assails me is in the right, and I who strike him down am also in the right. I defend against him not my right, but myself."12

I can also ask why, if I am going to prescribe that much, I should not more ambitiously prescribe that no one interfere with whatever particular purposes I may happen to have. I want the success of my particular projects, of course, as much as anything else, and I want other people not to interfere with them. Indeed, my need for basic freedom was itself derived from that kind of want. But the argument is certainly not going to allow me to prescribe for all my particular wants.

The argument depends on a particular conception of the business of making rules, a conception that lies at the heart of the Kantian enterprise. If I were in a position to make any rules I liked and to enforce them as an instrument of oppression, then I could make a law that suited my interests and attacked the competing interests of others. No one else would have a reason to obey such a law, except the reason I gave him. But the laws we are considering in these arguments are not that kind of law, have no external sanction, and respond to no inequalities between the parties. They are notional laws. The question "what law could I make?" then becomes "what law could I make that I could reasonably expect others to accept?" When we reflect on the fact that everyone asks it from an equal position of powerlessness — since these are laws for a kingdom where power is not an issue — we see that the question could equally be "what law could I accept?" and so, finally, "what laws should there be?"

If this is the question, asked in such a spirit, for such a kingdom, then we can see why its answer should be on the lines of Kant's fundamental principle of action, the Categorical Imperative of morality, which (in its first formulation) requires you to "act only on that maxim through which you can at the same time will that it should become a universal law." But the problem immediately becomes: Why should one adopt such a picture? Why should I think of myself as a legislator and — since there is no distinction — at the same time a citizen of a republic governed by these notional laws? This remains a daunting problem, even if one is already within ethical life and is considering how to think about it. But it is a still more daunting problem when this view of things is being demanded of any rational agent. The argument needs to tell us what it is about rational agents that requires them to form this conception of themselves as, so to speak, abstract citizens.

It might be thought that the question answers itself because, simply as rational agents, there is nothing else for them to be, and there is no difference among them. But to arrive at the model in this way would be utterly unpersuasive. We are concerned with what any given person, however powerful or effective he may be, should reasonably do as a rational agent, and this is not the same thing as what he would reasonably do if he were a rational agent and no more. Indeed, that equation is unintelligible, since there is no way of being a rational agent and no more. A more sensible test would be to ask what people should reasonably do if they did not know anything about themselves except that they were rational agents; or, again, what people should do if they knew more than that, but not their own particular powers and position.14 This is an interesting test for some things; in particular, it is a possible test for justice,
and in that role it can be proposed to those with a concern for justice. But it is not a persuasive test for what you should reasonably do if you are not already concerned with justice. Unless you are already disposed to take an impartial or moral point of view, you will see as highly unreasonable the proposal that the way to decide what to do is to ask what rules you would make if you had none of your actual advantages, or did not know what they were.

The Kantian project, if it is to have any hope, has to start farther back. It has to be, in a vital way, more like Kant's own project than the argument I have just outlined. The argument started from what rational agents need, and while what it said about that was true, it was not enough to lead each agent into morality. Kant started from what in his view rational agents essentially were. He thought that the moral agent was, in a sense, a rational agent and no more, and he presented as essential to his account of morality a particular metaphysical conception of the agent, according to which the self of moral agency is what he called a "noumenal" self, outside time and causality, and thus distinct from the concrete, empirically determined person that one usually takes oneself to be. This transcendental idea of the self, Kant believed, will be uncovered if we reflect on the requirements of freedom, requirements lying deeper than any that have been uncovered at the level of inquiry we have been pursuing up to now. He did not believe that we could fully understand this conception, but we could see that it was possible and could know that it was involved in both morality and rational action.

Kant's account presents great difficulties and obscurities. First, he believed that all actions except those of moral principle were to be explained not only deterministically but in terms of egoistic hedonism. Only in acting from moral principle could we escape from being causally determined by the drive for pleasure, like animals; and sometimes he marked this by saying that only actions of principle counted as exercises of the will (which he equated with practical reason) and hence were truly free. Our other actions, according to this way of putting it, are the product merely of causality — of "blind" causality, as people tend to say, unhappily.

since, as Kant himself recognized, such a causality can often enable agents, and certainly animals, to see very well where they are going.

I shall not go into the question of how far Kant's own theory can be rescued from these difficulties. Any theory that is going to provide foundations will certainly need to avoid them. We are interested in the idea that ethical considerations are presupposed by rational freedom, and this will have to mean a freedom to which the moral skeptic, among others, is already committed. It is open to Kant or another arguing like him to say that the moral skeptic is committed, in his desire for individual autonomy and rationality, to conceptions that are fully realized only in the moral law, but it will be useless to say that the moral skeptic must aspire to a kind of rational freedom quite different from anything manifested in non-moral practical intelligence or deliberation. The skeptic's commitment to freedom and rationality cannot be so detached from things he already experiences, such as the difference between deciding clear-headedly and finding himself doing things he did not intend. Moreover, this is not simply a dialectical point, about the hold one can hope to get on the skeptic. It is also a question of what conception of rational freedom it is reasonable to hold.

What we are looking for, then, is an argument that will travel far enough into Kant's territory to bring back the essential conclusion that a rational agent's most basic interests must coincide with those given in a conception of himself as a citizen legislator of a notional republic; but does not bring back the more extravagant metaphysical luggage of the noumenal self. The argument might go something like this. We have already agreed that the rational agent is committed to being free, and we have said something about what is required for that freedom. But we have not yet reached a deep enough understanding of what that freedom must be. The idea of a rational agent is not simply the third-personal idea of a creature whose behavior is to be explained in terms of beliefs and desires. A rational agent acts on reasons, and this goes beyond his acting in accordance with some regularity or law, even one that refers to beliefs and desires. If he acts on reasons, then he must not only be an agent but reflect on himself as an agent, and this involves his seeing himself as one agent among others. So he stands back from his own
desires and interests, and sees them from a standpoint that is not that of his desires and interests. Nor is it the standpoint of anyone else's desires and interests. That is the standpoint of impartiality. So it is appropriate for the rational agent, with his aspiration to be genuinely free and rational, to see himself as making rules that will harmonize the interests of all rational agents.

In assessing this line of argument, it is important to bear in mind that the kind of rational freedom introduced by it is manifested, according to Kant, not only in decisions to act but also in theoretical deliberation, thought about what is true. It is not merely freedom as an agent — the fact (roughly speaking) that what I do depends on what I decide — that leads to the impartial position, but my reflective freedom as a thinker, and this applies also to the case of factual thought. In both cases, Kant supposed, I am not merely caused to arrive at a conclusion: I can stand back from my thoughts and experiences, and what otherwise would merely have been a cause becomes a consideration for me. In the case of arriving by reflection at a belief, the sort of item that will be transmuted in this way will be a piece of evidence, or what I take to be evidence: it might for instance be a perception. In the case of practical deliberation, the item is likely to be a desire, a desire which I take into consideration in deciding what to do. In standing back from evidence, or from my desires, so that they become considerations in the light of which I arrive at a conclusion, I exercise in both cases my rational freedom. When, in the practical case, I adopt the standpoint outside my desires and projects, I may endorse my original desires, as in the factual case I may endorse my original disposition to believe.

The fact that Kant's account of rational freedom is meant to apply to factual deliberation as much as to practical brings out what is wrong with the Kantian argument. What it says about reflection does indeed apply to factual deliberation, but it does so because factual deliberation is not essentially first-personal. It fails to apply to practical deliberation, and to impose a necessary impartiality on it, because practical deliberation is first-personal, radically so, and involves an I that must be more intimately the I of my desires than this account allows.

When I think about the world and try to decide the truth about it, I think about the world, and I make statements, or ask questions, which are about it and not about me. I ask, for instance,

Is strontium a metal?

or confidently say to myself

Wagner never met Verdi.

Those questions and assertions have first-personal shadows, such as

I wonder whether strontium is a metal,

or

I believe that Wagner never met Verdi.

But these are derivative, merely reflexive counterparts to the thoughts that do not mention me. I occur in them, so to speak, only in the role of one who has this thought.

Of course, I can occur in my own thoughts in a more substantive and individual way. My thoughts may be specifically about myself, as in

Am I ill?

Thoughts of that kind are about myself in a sense in which other thoughts I have are not about myself, but about someone or something else. More interestingly, I may occur in my thought as a locus of evidence, as in

It looks blue to me.

In such a case I occur as specifically myself, and my actual psychological properties are relevant (thus, given my eyesight, the thing's
looking blue to me may be a reliable indicator of its being green). If I ask

What do I think about this question?
in one sense it also involves a specific reference to myself, with my actual psychological properties; it can be an invitation to me to find out about my beliefs, as I might find out about someone else’s (if not in exactly the same ways). But

What should I think about this question?

where that has the same effect as

What is the truth about this question?
is again a case in which I occurs only derivatively: the last question is the primary one.

Because of this, the I of this kind is also impersonal. The question,

What should I think about this question?
could as well be

What should anyone think about this question?

This is so, even when it means

What should I think about this on the evidence I have?

This must ask what anyone should think about it on that evidence. Equally, what anyone truly believes must be consistent with what others truly believe, and anyone deliberating about the truth is committed, by the nature of the process, to the aim of a consistent set of beliefs, one’s own and others.19

It is different with deliberation for action. Practical deliberation is in every case first-personal, and the first person is not derivative or naturally replaced by anyone. The action I decide on will be mine, and (on the lines of what was said earlier about the aims of action) its being mine means not just that it will be arrived at by this deliberation, but that it will involve changes in the world of which I shall be empirically the cause, and of which these desires

and this deliberation itself will be, in some part, the cause. It is true that I can stand back from my desires and reflect on them, and this possibility can indeed be seen as part of the rational freedom at which any rational agent aims. This goes somewhat beyond the considerations about freedom and intentionality acknowledged earlier in the discussion, but it still does not give the required result in relation to morality. The I of the reflective practical deliberation is not required to take the result of anyone else’s properly conducted deliberation as a datum, nor be committed from the outset to a harmony of everyone’s deliberations — that is to say, to making a rule from a standpoint of equality. Reflective deliberation about the truth indeed brings in a standpoint that is impartial and seeks harmony, but this is because it seeks truth, not because it is reflective deliberation, and those features will not be shared by deliberation about what to do simply because it too is reflective. The I that stands back in rational reflection from my desires is still the I that has those desires and will, empirically and concretely, act; and it is not, simply by standing back in reflection, converted into a being whose fundamental interest lies in the harmony of all interests. It cannot, just by taking this step, acquire the motivations of justice.

Indeed, it is rather hard to explain why the reflective self, if it is conceived as uncommitted to all particular desires, should have a concern that any of them be satisfied. The reflective self of theoretical or factual deliberation has a unity of interest with prereflective belief: each in its own aims at truth, and this is why the prereflective disposition to believe yields so easily, in the standard case, to corrective reflection. But on the model we are considering there is not an identity of interest between the reflective practical self and any particular desires, my own or others. It is unclear, then, why the reflective self should try to provide for the satisfaction of those desires. This is just another aspect of the mistake that lies in equating, as this argument does, reflection and detachment.

Some deep questions remain about what it is to take the impartial perspective if one does possess the motivations of justice. How can an I that has taken on the perspective of impartiality be left with enough identity to live a life that respects its own interests? If
morality is possible at all, does it leave anyone in particular for me to be? These are important questions about both morality and life: about morality because, as a particular view of the ethical, it raises that question in a particularly acute form, and about life because there are, on any view of ethical questions, real issues about the relations between impartiality and personal satisfactions and aims — or, indeed, personal commitments that are not necessarily egoistic but are narrower than those imposed by a universal concern or respect for rights. Some of these questions will arise later. They concern what happens to personal desire and deliberation under the influence of the impartial standpoint, to the extent that one achieves it. What has been shown in this chapter, I believe, is that there is no route to the impartial standpoint from rational deliberation alone.

The impartial standpoint can be called upon for a different purpose, not to argue someone all the way from bare practical reason to the concerns of justice or benevolence, but to support or demand some ethical conceptions rather than others. The question now is: Given people who are in some general sense committed to thinking in ethical terms, how should they think? Are their ethical thoughts sound?

I am not concerned here with every kind of critique of existing ethical attitudes and beliefs. There are many styles of critique, and the most potent of them rely, as they always have, not so much on philosophical arguments as on showing up those attitudes as resting on myths, falsehoods about what people are like. Even among the criticisms that involve more distinctively philosophical argument, not all of them are my immediate concern. Some of these patterns of argument serve in a local fashion, to bring out the consequences of ethical positions or to convict them of incoherence. They are instruments of ethical argument. In this chapter and the next, I am concerned with a more elaborate, thoroughgoing, and ambitious kind of structure, the ethical theory. (Later I shall consider the idea that some of the instruments of ethical argument, thoroughly applied, are themselves enough to generate ethical theory.)

What is an ethical theory? The most helpful use of that expression
can best be caught by a rather complex definition. An ethical theory is a theoretical account of what ethical thought and practice are, which account either implies a general test for the correctness of basic ethical beliefs and principles or else implies that there cannot be such a test. It is the first kind of ethical theory, the positive kind, that concerns me here. In this chapter I shall give an account of two leading styles of positive ethical theory, and in the next I shall consider the deeper motivations for theories of this kind and their relation to practice. First, however, I must say something about the negative kind of ethical theory, and about the point of putting the definition in this peculiar way.

It does not matter much how the expression “ethical theory” is used, so long as one’s use is made clear. There is a reason, though, for using it in the way suggested by this definition, and it involves a significant philosophical point. Twenty or thirty years ago, it was standard practice to distinguish “ethical” from “meta-ethical” theories. The first made substantive claims about what one should do, how one should live, what was worthwhile, and so on. The second concerned itself with the status of those claims: whether they could be knowledge, how they could be validated, whether they were (and in what sense) objective, and so on. The idea that lay behind the distinction was, naturally enough, that the two types could be taken apart and that a theory of the meta-ethical sort would not, as such, have ethical implications.

It is important to separate this proposed distinction from two other ideas that have often been associated with it. One is that the meta-ethical study should be linguistic, an inquiry into the terms used in ethical discourse. This involves an additional view, about the nature of philosophy — one that has not been very fruitful in moral philosophy. Although the distinction can be separated from this idea, the linguistic formulation probably helped to encourage it, because of a general assumption that it must be possible to distinguish between the means a language provides for saying things and the particular things that happen to be said in it. This assumption was widely made at that time and issued in other distinctions, such as that between analytic and synthetic, which are now also regarded with less favor. But even if the preoccupations of linguistic philosophy encouraged the distinction between the ethical and the meta-ethical, that distinction is not committed to a linguistic formulation.

Another idea that is separable from the distinction, and indeed separable from it even when it takes a linguistic form, is that philosophy should not contain any ethical assertions and should confine itself to the meta-ethical. This policy obviously rests on further assumptions, once more about the nature of philosophy, and they have not universally been made. Thus Moore, whose Principia Ethica made an emphatic and influential distinction between saying what goodness is and saying what things are good, allowed himself in that book to try the second as well as the first, and while it was the distinction that was influential with philosophers, it was his account of what things are intrinsically good that impressed others. Moore did, however, have a view about goodness and our knowledge of it, in terms of which it was to some extent appropriate to philosophy that it should try to say what things are good. He thought that goodness was detected by a kind of intellectual discrimination, and part of that process at least (it was too poorly defined to make it clear how much) was enough like intellectual analysis to make it intelligible that philosophy, or the capacities of philosophers, might have something to contribute to it. But someone who thinks that the business of philosophy is primarily analysis, and that what is involved in making substantive ethical judgments is quite different from intellectual analysis, will see no reason why those judgments should be part of philosophy, and will try, as some philosophers did twenty or thirty years ago, to leave them out of it.

The distinction between the ethical and the meta-ethical is no longer found so convincing or important. There are several reasons for this, but the most relevant here is that it is now obvious (once again obvious) that what one thinks about the subject matter of ethical thought, what one supposes it to be about, must itself affect what tests for acceptability or coherence are appropriate to it; and the use of those tests must affect any substantive ethical results. Conversely, the use of certain tests and patterns of argument can imply one rather than another view of what ethical thought is. A
theory that combines views on what ethical thought is and how it
should be conducted, with substantive consequences of conduct­
ing it in that way, is a positive ethical theory.

Some views about the content and nature of ethics, however, imply that there are no tests. The most extreme of these views says that holding an ethical position simply consists of choosing one and sticking to it. There seems to me good reason to call that an ethical theory too, a negative one. But this should be distinguished from a theory about the nature of ethical thought that leaves open the question whether there could be such tests. One may be fairly convinced and definite about the account to be given of the ethical, and remain skeptical about the chances of there being these tests; and there are options more complex than that, according to which there may be tests in some cultural circumstances and not in others. That is the kind of account I give in this book, and there is point in not calling it an ethical theory. Ethical theories are philosophical undertakings and commit themselves to the view that philosophy can determine, either positively or negatively, how we should think in ethics. It is this negative option that philosophers usually had in mind when in the past they said that philosophy could not determine how we should think in ethics. In contrast, I want to say that we can think in ethics, and in all sorts of ways, unless our historical and cultural circumstances have made it impossible—but that philosophy can do little to determine how we should do so. The purpose of using “ethical theory” in the way I suggest is to bring out the similarity of the positive and the negative theories in the claims they implicitly make for philosophy. It may at this stage of the argument seem a fine point, but I hope that by the end it will not seem so. The aim is to reach an outlook different from that of any of these theories. It is an outlook that embodies a skepticism about philosophical ethics, but a skepti­cism that is more about philosophy than it is about ethics.

We must now turn to positive ethical theories. There are several kinds of ethical theory, and there are several ways of classifying them, which yield different kinds of kinds. No classification is uniquely illuminating, but one helpful distinction is that between two basic styles, the contractual and the utilitarian. The central idea of contractualism has been formulated by T. M. Scanlon, in relation to its account of moral wrongness: “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 no­one could reasonably reject as a basis for informed, unforced, gen­eral agreement.” (Scanlon and other writers I shall be discussing usually speak of morality; I shall sometimes do the same.) This account of wrongness goes with a particular theory of what moral thought is about, or of what ultimate moral facts there are. On this theory, moral thought is concerned with what agreements people could make in these favored circumstances, in which no one was ignorant or coerced. The theory also gives an account of moral motivation. The basic moral motive is “a desire to be able to justify one’s actions to others on grounds they could not reasonably re­ject.” It can be seen how close this complex of ideas is to the Kantian conceptions discussed in the last chapter. Now, however, it is not a matter of trying to show that every rational agent must be a citizen legislator of a notional republic. It is a question of what rules would be acceptable to people who are assumed to be already inter­ested in reaching agreement.

Utilitarianism, by contrast, takes facts of individual welfare as the basic subject matter of ethical thought. There are many species of utilitarianism. They disagree about how welfare is to be assessed, and about other questions: whether, for instance, it is the individ­ual act that should be justified in terms of maximizing welfare, or instead some rule, practice, or institution. (This is the difference between direct and indirect utilitarianism.) All the variants agree on aggregating welfare, that is to say, adding together in some way the welfare of all the individuals involved (this formula, even the word “involved,” raises many difficulties).

I shall look at these styles of ethical theory in greater detail. But even these introductory sketches give some idea of how they might lead to different results. One difference lies in the constituency of morality as it is most naturally defined by the theories: that is, those with whom the system is in the first place concerned. The natural
constituency for contractualism consists of those to whom you could conceivably try to justify your actions — in the simplest interpretation, other moral agents. This can be extended to a concern for the interests of others who are unable to give or receive justifications — small children, for instance, or the mentally handicapped. In such cases we naturally think, as we do in the law, of trustees acting on those people’s behalf. By a further extension, animals may also receive consideration, but they are farther away from the primary constituency. We would expect contractualism to give an account of concern for animals that is different from that given of moral relations between people. The idea of a contract, even in this minimal and schematic form, always brings in as its first concern equal relations between agents who are both the subjects and the objects of moral thought.

Utilitarianism looks in a different direction. One of the most natural interpretations of the welfare with which it is concerned (and historically the earliest) is pleasure and the absence of pain, and the natural utilitarian constituency consists of all creatures capable of feeling pleasure and pain. This basis has been refined by modern work, and the constituency is now likely to be defined in terms of those who have preferences or wants, and can suffer from the frustration of those wants. In most versions, this still includes animals in the primary constituency: in fact, it includes some animals more naturally than some humans (moribund humans, for instance). This conception appeals to one moral motivation, benevolence. At the same time, it introduces a disparity between moral agents, on the one hand, and beneficiaries of morality on the other, the second class being, right from the beginning, larger than the first. This feature of utilitarianism comes from its welfarism.

It has another important feature, which comes from its being a kind of consequentialism and judging actions in terms of their consequences. Any form of consequentialism locates ethical value ultimately in states of affairs. (In the case of utilitarianism, which is welfarist consequentialism, that value is found in differences of welfare located in states of affairs.) This has the result that, for utilitarianism, agency comes in only secondarily: our basic ethical relation to the world, as agents, is that of being the cause of desirable or undesirable states of affairs. Our basic ethical concern is to bring it about, so far as we can, that there is more welfare or utility in the world rather than less, and, in the simplest version of utilitarianism, we should simply act in the most efficient way to bring that about. It is a question of what causal levers are at that moment within reach. Sometimes the causal connections through which I can affect outcomes run through other people’s actions, but this makes no special difference. It is simply a matter of what changes produce most welfare. This means that there are states of affairs I can affect with respect to welfare which, because I can do so, turn out to be my concern when, on nonutilitarian assumptions, they would be someone else’s concern. Moreover, because the class of beneficiaries is larger than that of agents, there are situations that turn out to be someone’s concern when on nonutilitarian assumptions they would have been no one’s concern.

These considerations bring out another difference between utilitarianism and contractualism — once more, on the most immediate and natural interpretations of those theories. The demands of utilitarianism for maximum welfare production are boundless. There is no limit to what a given person might be doing to improve the world, except the limits of time and strength. Moreover, because the relations of possible states of affairs to any given person’s actions are indeterminate, the demands are boundless in the further sense that there are often no clear boundaries between the demands on me and the demands on someone else. Utilitarian theorists go on (with varying degrees of enthusiasm) to put back restrictions on what a given individual may be required to do, saying that it is usually more efficient if you care specially about your own children, for instance, or if you relax occasionally from good works. Contractualists, and many others who are not theorists of either sort, will complain that this gets things back to front: my rights to my own children and my own time are not theirs to give back.

The idea that there should be a limit to “the strains of commitment,” as John Rawls puts it, is one thing that helps to form his theory, the richest and most complex contractual account of ethics.
yet advanced. Rawls's theory of justice aims to find principles to govern social and political life rather than individual conduct. But it starts from a moral basis, and also has important consequences for purely moral thought.

Rawls's theory is an elaboration of a simple idea: a fair system of arrangements is one that the parties can agree to without knowing how it will benefit them personally. This is worked up into a fiction of an Original Position in which people choose social principles behind a "veil of ignorance," which conceals from them their own prospective social positions and indeed their individual tastes and interests. It does not conceal general propositions, such as the findings of the social sciences, so they have some information to work on, but no information that enables any of them to discriminate in his or her own favor. The ignorance is thus less radical than that implied in the last chapter, where I touched on the Kantian idea of people choosing as rational agents "and no more." (The model has the revealing consequence that Rawls must assume, implausibly, that knowledge of history is not essential to social scientific understanding — unless he allows, even less plausibly, that you can know the course of history without knowing your place in it. The fact that the theory, like other such theories, is radically ahistorical is an important fact about it.) The parties have to make, in these circumstances, a self-interested choice of social arrangements. They are not in a good position to do it, since they do not know what selves they are, but it does mean that they do not include any benevolent or altruistic principles in the basis of their choice. Rawls is not to be interpreted here as trying to move to social justice from personal self-interest. The point is that a self-interested choice in ignorance of one's identity is supposed to model in important respects non-self-interested or moral choice under ordinary conditions of knowledge.

The result of the deliberation in the Original Position is that the parties accept two fundamental principles of justice:

1. Each person is to have an equal right to the most extensive liberty compatible with a similar liberty for others.
2. Social and economic inequalities are to be arranged so that they are both (a) to the greatest benefit of the least advantaged and (b) attached to offices and positions open to all under conditions of fair equality of opportunity.

The second of these principles is based on the idea that parties will use in their deliberations a "maximin" rule, a rule that ranks alternatives by their worst possible outcomes. It is a distinctive feature of Rawls's theory. The choice of this principle in the Original Position is supposed to rest not on any peculiarly conservative bias of the parties (who cannot allow for a special taste in that or any other direction), but on the peculiar character of the choice, which consists in the fact that the parties have no probabilities available to them; that they have no very great interest in benefits over the minimum; and that the worst outcomes involve "grave risks" that one could not accept.

This expresses some important ideas about fundamental goods — for instance, that slavery is simply unacceptable, whatever benefits it might confer. Indeed, as Rawls's rejection of all probability calculations shows, he is committed to the conclusion that even a society with a very small number of slaves would be unacceptable. This may be a welcome consequence of a moral theory of justice, but it does not follow naturally from the model of rational choice under ignorance. If self-interested rational choice is what is at issue, it is hard to see how the question of probabilities can altogether be avoided, or how, if the probability of ending up as a slave were small enough, it would not be rational for the parties to choose a system involving slavery if it conveyed large enough other benefits. For reasons of this kind, the decision-theoretical or rational-choice element in Rawls's model has been much criticized.

There is also an important question of what the goods are, in terms of which the parties are supposed to make their rational choice. Rational-choice theory, in its ordinary uses, normally works on a basis of utility or individual welfare, and this is a function of the agents' preferences and tastes (we shall see later that in the hands of the utilitarians welfare does not remain such a simple matter). But the contracting parties do not have any known individual preferences or tastes, so this is not available to them. Rawls makes
his parties choose by reference to a list of "primary goods," which are liberty and opportunity, income and wealth, and the bases of self-respect. These goods are given by what he calls "the thin theory of the good"; the idea is that these are goods that everyone is going to want if they want anything. But the question is more complex than this. The list of primary goods does not plausibly look as if it had been assembled simply from the consideration that they are uniquely necessary for pursuing anything. From that consideration we are not likely to derive more than liberty. It is hard to see, also, how the parties could avoid the reflection (available to them from their knowledge of general social facts) that some of these primary goods, notably money, are more important in some societies than in others.

Why are primary goods introduced at all? It is not technically impossible for parties without any known particular preferences to choose one social state over another. We could suppose them to choose a situation in which they (that is to say, anyone they might turn out to be) get more of what they prefer in that situation than they do of what they prefer in other situations. As we shall see, this is what is done by R. M. Hare in constructing a utilitarian theory. Such comparisons may indeed be fanciful, but that is not Rawls's objection to their use in the Original Position. Rather, he has refused to let his parties think merely in terms of what they would prefer in the various social situations. His parties are reluctant, as we actually are, to count a situation as acceptable just because they would find it acceptable if they were in it (we noticed our own reluctance earlier, with "real interests"). So even if the social sciences told us, as in fact they are very unlikely to tell us, that most slaves are content to be free from freedom, this would not give the parties a reason to choose slavery: from the standpoint of free people freely choosing a polity, this is not an option. Rawls has the right, indeed is right, to carry such convictions into his ethical theory, but they are not best represented by the machinery of rational choice deployed under selective conditions of ignorance. The primary goods may perhaps be better seen in terms of a fundamental ethical conception of the person, and Rawls himself has now moved in this direction.

Formally speaking, utilitarianism is itself an option under contractualism. If the contractualist question is asked not about particular principles or practices, but about an entire set of principles, as it is by Rawls, then the parties could choose some utilitarian system as the answer. Granted the kind of differences we have already noticed in the typical outcomes of the two ways of thinking, this is unlikely, but it is not excluded by the machinery. Consider some people asked to select a set of principles. They are armed with utilities (instead of an index of primary goods), and there are no further restrictions on their choice, except that they are subject to the one dimension of ignorance, that they do not know who they will be in the world governed by the principles they select. The idea that certain principles are impartially acceptable is then equated with their being those principles that would be selected by someone who believed that he had an equal chance of being anyone in the outcome. (It can now be seen that there need be only one person choosing.) This is the approach of John Harsanyi, who argues that it yields a set of principles that would maximize the average utility of the affected people. This resembles a contractualist argument, but it has a utilitarian outcome.

In its most familiar versions, however, utilitarianism starts from ideas of welfare, or of people's interests. Its project, which (in its simplest form) consists of considering everyone's welfare under various alternative outcomes and compounding it, involves serious technical difficulties as well as deep conceptual ones. One of them is that except in the simplest cases the set of people affected by various outcomes will not be the same, and people who have to be considered under one alternative may well not exist under others. But I shall not pursue technical difficulties here. I said earlier that for utilitarianism the characteristic moral motive was benevolence. That term is vague, and it can also be misleading, particularly if it suggests warm feelings of personal attachment or, again, any kind of sentiment one naturally feels in greater degree for some people than others. Utilitarian benevolence involves no particular attachments, and it is immune to the inverse square law. The term stands for a positive relation to other people's desires and satisfactions, which the benevolent person has only
because they are the desires and satisfactions of others. This rough idea needs work before it can play a part in ethical theory, and the work is done in different ways by different theorists. In considering the important question of utilitarianism's impartial attitude to desire satisfaction, I shall discuss an interesting version of utilitarianism developed by R. M. Hare, who treats the agent's relation to others' desires (what I vaguely called "benevolence") in terms of imaginative identification.

Hare's theory starts from some claims about the nature of moral judgments: they are prescriptive, and they are universal. "Prescriptive" is a term relating to language (I have touched briefly on a possible use of it in Chapter 4 and shall have more to say about it in Chapter 7). A prescriptive utterance is of the type "let so-and-so be done," and Hare takes such an utterance, if sincere, to express a desire or preference. Moreover, every preference can be expressed in a prescription; so any agent who has preferences is in a position to make prescriptions. Yet the agent is not yet committed to making universal prescriptions — those come in with moral language, in particular, moral uses of ought. Hare thus does not make the claim that the presuppositions of any practical reasoning involve a universal prescription. So far as that goes, Hare and I are in agreement, that the commitments of moral reasoning can be avoided by not engaging in moral reasoning.

The effect of making a universal prescription, in judging that I ought to do a certain thing, is that I accept that anyone else ought to act similarly in similar circumstances. In particular, I accept that this ought to be the case if I were at the receiving end of the action. In considering what I ought to do, therefore, I must consider what it would be like to be the other people affected; in doing this, I apply a "role-reversal test" and think what I would want or prefer if I were in their positions. I should, if thinking ideally, conduct this thought experiment with regard to every person, or indeed every sentient creature, involved in similar situations.

The use of a role-reversal test is not peculiar to utilitarianism. In one form or another it is a basic item of ethical thinking, and a version of it is involved in Kant's Categorical Imperative. The distinctively utilitarian results follow from the special and radical interpretation that Hare's theory gives to the idea of thinking oneself into someone else's position. On this interpretation, an agent will have realized what he would prefer if he were in that position only if he now acquires a corresponding actual preference that applies to the hypothetical situation. I shall come later to Hare's reasons for this idea and to criticism of it. First, however, we should see what a crucial part it plays in the transition to utilitarianism. If anyone carried out an ideally complete thought experiment under these requirements, he would actually acquire preferences that would correspond to every preference held by anyone who was affected by the situation. All the preferences would thus be agglomerated into one individual. How can the agent decide, given this agglomeration? These are now his preferences, and he can bring to them certain rational requirements that supposedly apply to any first-personal deliberation. But he cannot on ethical grounds discount or downgrade any preferences he has acquired from identification with others — nor, come to that, any of those he started with — since this ideal level of reflection is supposed to criticize all ethical grounds, and none of them can be taken for granted. All that the ideally reflective agent is given, at the ethical level, is the process of additive identification itself. So once the preferences have been adjusted in the light of rational criteria that apply even to first-personal deliberation, there is nothing to do except compare their relative strengths and choose between the various outcomes on the basis of that comparison. The result is utilitarianism.

This structure is equivalent to one version of what has been called Ideal Observer theory. This postulates one omniscient, impartial, and benevolent observer — he might be called the World Agent — who acquires everybody's preferences and puts them together. The test of what should be done (or, in indirect versions of the theory, of what practices or institutions should be adopted) then becomes what would be chosen by such an observer. As Hare says, his own model comes to the same thing. There is another version of Ideal Observer theory, which leaves out the condition of benevolence and does not imagine the observer actually to take on everyone's preferences. As Roderick Firth put it (perhaps a little
quaintly) in a well-known exposition of the theory, the observer is "omniscient, disinterested, dispassionate, but otherwise normal." This version of the theory, in which the various preferences are not aggregated into a World Agent but merely surveyed from outside in a dispassionate spirit, is not supposed to lead necessarily to utilitarianism. Utilitarianism is merely one candidate that might itself be selected by the theory's test. In this form, however, Ideal Observer theory faces the objection that if the observer is not given some motivation in addition to his impartiality, there is no reason why he should choose anything at all; and unless that motivation is benevolent—or positively related to the preferences he knows about—he might as well choose to frustrate as many preferences as possible.

In earlier work Hare did not make the hypothetical identifications so complete, with the result that he was left with what he saw as a problem, the possibility of a "fanatic" who was so wedded to certain ideals that he would accept the hypothetically unpleasant results of the role-reversal argument. Thus a convinced Nazi might accept the prescription "let me be killed if I were a Jew." Hare now regards this problem as having been overcome, by the same process of thought that led him to a utilitarian position. The Nazi, if he engages in the ideal process of thinking, will, in identifying with Jews, lose in relation to the hypothetical situations his antisemitic preferences and will, for each Jew, acquire an actual preference against antisemitism. This process of thought will itself constitute a utilitarian critique (assuming that the sums come out right) of antisemitism. It is of course accepted that an actual racist may refuse to engage in this process of thought, but that does not undermine the argument.

It is striking how strong the claims are that Hare makes for the powers of rational argument in ethics. Why was the fanatic regarded as a problem in the first place? Hare says, "It would expose a gap in the defences of utilitarianism if a fanatic could listen to, and understand, all the arguments, and admit all the facts, adduced by a utilitarian, and still sustain his opinion." At first glance, this seems to imply a super-power view of defense, that you are adequately defended only if you can annihilate the other side. As against this, it is surely possible that you might hold a rational, or reasonable, set of ethical beliefs, and yet there be other people who held different ethical beliefs that you might indeed deplore but could not demonstrate to be inconsistent or factually wrong. (Hare himself, in his earliest work, The Language of Morals, did believe this.) More interestingly, you may think that some positions that differ from your own are indeed irrational, and racism is one of them; but that their irrationality cannot necessarily be shown by the same arguments in every case, nor by some pattern of argument central to building your own ethical beliefs. There may be something specially irrational about racism.

Hare's equation of defense and attack comes from two sources, one rather special to his outlook, the other shared by many ethical theories. The general point is that ethical theories in this style can readily be seen as offensive weapons, aimed against prejudice, so that if there is an important style of prejudice that is immune to them, they are not well designed for their job and are likely to replaced, if not by prejudice, then by an ethical theory with more firepower. To some extent, this is true of all of them, although they differ from one another in their aggressive ambitions.

Hare's special reason for his view of his defenses is that he believes his argument for utilitarianism to follow strictly from the meaning of moral words. It will indeed be a gap in the defenses of that claim if there can be someone who correctly uses moral language but consistently refuses the theory: that claim will be not only defenseless but defeated. But the claim is unreasonable. Alternative theories cannot plausibly be shown to misuse or misinterpret moral language. Even if there were one basic characteristic of "moral language" as such, and even if that lay in its being prescriptive and universal, this would still not lead inescapably to the theory. There are other interpretations of what it is to accept a prescription, and of what counts as universalizability, that would lead to different theories. As John Mackie argued, there are various degrees of universalizability less extreme than the ultimate stage represented by Hare's theory, which sinks all the agent's tastes and ideals into the thought experiment of identification. In Mackie's words, "it is at most the first stage, the ruling out of purely numeri-
cal differences as morally irrelevant, that is built into the meaning of moral language.\(^{16}\)

In everyday uses of role-reversal arguments, less ambitious than the attempt to found all moral considerations on them, it is often natural to include personal tastes in the imagined identification, but to leave out ideals or ethical beliefs. Making this distinction involves resisting a characteristic move of utilitarian thought, which has been called reduction and defined as “the device of regarding all interests, ideas, aspirations and desires as on the same level, and all representable as preferences, of different degrees of intensity, perhaps, but otherwise to be treated alike.”\(^{17}\) Utilitarian writers often start with a plea for equal consideration of everyone’s interests, and then extend this upward (so to speak) to ideals and downward to mere tastes. This assimilation gets things out of proportion in more than one direction. In one way it underestimates the significance of ideals or ethical conceptions, and requires an agent to abandon any stand of principle or deeply held conviction if a large enough aggregate of preferences, of whatever kind, favors a contrary action. The assimilation does not give our convictions enough weight in our own calculations. At the same time, it can give other people’s convictions too much weight. While Hare’s thought experiments give an argument against the racist fanatic, they do not give the right sort of argument.

There is, first of all, the point that the sums have to come out right. It is naturally characteristic of utilitarianism that its results depend on calculations but, in connections such as this, that feature is particularly undesirable. If racist prejudice is directed toward a small minority by a majority that gets enough satisfaction from it, it could begin to be touch and go whether racism might not be justified. The point is not how likely that is to arise, or in what circumstances, but that the whole question of how many racists are involved cannot begin to be an acceptable consideration on the question whether racism is acceptable (contrast Rawls’s treatment of slavery). Moreover, on the utilitarian argument it emerges as a consideration — though, if the sums come out right, not a decisive or winning consideration — that racists get some satisfaction out of the sufferings of the Jews; but this cannot be a consideration at all. This does not mean that the sufferings of racists never count. It means that sufferings they experience solely because of their racist opinions do not count. Harsanyi, indeed, has built into his system a provision that aims to deal with this kind of problem, by rather briskly excluding “antisocial” preferences from the count. But he does not explain how they are to be defined; his rationale for the provision suggests that what is in question are antiutilitarian preferences.\(^{18}\)

I have already mentioned the requirement of Hare’s theory, that the preferences, when they have all been gathered into one agent, should be modified by reference to criteria of first-personal rationality. What are eventually taken into account are not necessarily the actual preferences of agents (including their actual hypothetical preferences — that is, the preferences they would as a matter of fact have in the hypothetical situations), but rather their “perfectly prudent preferences,” which is what they would prefer if they were fully informed and unconfused in their thinking. A similar provision is made by Harsanyi. Moral thinking has been assimilated (in different ways in the two theories) to prudential thinking by a single individual. What it is prudentially rational for an ordinary agent to do does not necessarily correspond to what he actually prefers, since he may be confused or misinformed. Our knowledge of our future preferences also comes into this. We should include among our now-for-then preferences (as Hare calls them in a helpful terminology) our anticipated then-for-then preferences; the way we do this, according to Hare, is that we take on actual now-for-then preferences as surrogates, exactly as the reflective agent takes on other people’s preferences.\(^{19}\) There are many complications involved here. Here I want only to bring out the kind of treatment that Hare’s model gives to these aggregated preferences.

This process of correcting preferences (their idealization, as we might say) is appropriate to models in which all preferences become notionally one person’s preferences, such as the World Agent interpretation of Ideal Observer theory. It is appropriate, though, only if the model is taken literally; and if it is taken literally, even to a slight
degree, it becomes clear how bizarre it is. Any one agent who had projects as conflicting, competitive, and diversely based as the World Agent's would be (to put it mildly) in bad shape. He would need a set of values or second-order desires to give some weighting to his array of preferences. But if the World Agent has any of those, and they are still recognizable in the aggregate of preferences, he once again has too many. The truth is that this aggregate of preferences is simply unintelligible unless they are understood to be the preferences of different people. The device of the World Agent requires us to forget that fact, to see the ethical world as a sea of preferences. So, in varying degrees, do all forms of utilitarianism. This idea is often criticized in terms of its ethical results, but the fundamental objection is that it makes no sense as an interpretation of the world. It is because of this that it makes no sense of ethics. “The separateness of persons,” as John Findlay put it, is “the basic fact for morals.”

If utilitarianism is interpreted in terms less drastic than those of the World Agent model, the idealization of preferences becomes less appropriate. Certainly it is never appropriate merely because an agent’s preferences are based on false information. It will be appropriate if, as a result of action from that preference, the agent and others get less utility than they would get from action based on a correction of preferences. To take a political example, if a utilitarian administration operates not on the basis of what people prefer, but on the basis of what they would prefer if they were better informed, it is possible that those for whom it acts will always be discontented with what is actually done, since they may never lose their errors and, if they do not, will never actually have the idealized preference the policy is designed to satisfy.

The doubtful role of idealization in utilitarian theory is connected with the process of reduction I have already mentioned, by which interests are assimilated to preferences. Idealization or correction is appropriate when one is thinking about people’s interests — one of the basic facts about people’s interests is that they can be mistaken about them. The question is not whether you can appropriately correct others’ preferences when thinking about their interests, but how far you have the right to act on the basis of those corrections if the people concerned do not recognize them. But if you are simply concerned with how much preference-satisfaction the world will contain, the question will be different: whether idealization in a given case will in the long run create more utility. It can be seen how these two different questions will naturally figure in two different conceptions of politics.

Idealization is ambiguously related to the role-reversal test itself. As we have seen, it can be appropriately applied in the World Agent’s deliberations, after the preferences have all become his. Yet this result conflicts with the spirit in which the thought experiments of identifying with others were recommended to us in the first place. The original question was “how would it be for me if I were in his position?” This is interpreted by Hare as equivalent to “how would it be for him?”, none of the original me is left over in the transfer. (The hypothetical is taken as it is in the well-known reply to the remark “If I had been Roosevelt, I would not have made all those concessions to Stalin”: “Don’t be silly, if you had been Roosevelt, you would have done whatever Roosevelt did.”) But if this is to be the degree of identification (total), then if another’s preferences are mistaken, the preferences I imagine myself into are equally mistaken, and if identification is the point, they should remain so. In the outcome, however, my total sympathetic identification with the other person issues in my improving his preferences. This is a compact illustration of a truth about all utilitarian politics, that benevolence gets credentials from sympathy and passes them on to paternalism.

At this point we should look back at Hare's reasons for the interpretation that makes each ideally reflective person into a version of the World Agent. Its roots can be found in the relations that Hare sets up between two propositions which, on the face of it, are very different: (1) I now prefer with strength S that if I were in that situation X should happen rather than not. (2) If I were in that situation, I would prefer with strength S that X should happen rather than not. “What I am claiming,” Hare writes, “is not that these propositions are identical, but that I cannot know that (2), and what that would be like, without (1) being true.” Hare's
claim, that is to say, is about knowledge: I cannot now know that in a certain situation I would prefer with a certain strength that X should happen, unless I now prefer with that same strength that X should happen in that situation.

This claim seems hard to accept even if the I of the hypothetical situation is straightforwardly me, as in cases of buying insurance and other such prudential decisions. I indeed know, for instance, that if my house caught fire, I would prefer, with the greatest possible intensity, that my family and I should get out of it. Since I am a moderately rational agent, I take some action now to make sure that we could do that if the situation arose, and that action comes of course from a preference I have now. But there is no sense at all in which that present prudential preference is of the same strength as the preference I would have if the house were actually on fire (driving almost every other consideration from my mind), and it is not rational that it should be. For one thing, its strength will be formed in part by the probability of the imagined situation.

In the case of the thought experiment that goes with the radical version of the role-reversal test, it might be said that the probability of the imagined situation is always zero, since it is the probability of being someone else. But like this, that is unfair, since the experiment relates to situations described in general terms, so it is a question of satisfying some general description, not of blankly being another individual. Nevertheless, the probability in many cases will still be zero; and in any case the probability is not supposed to figure in the argument. Granted this, it seems even less plausible that the derived preference should be of the same intensity as the preference that would exist if the situation came about. Indeed it is not even clear what it means to say that it is or that it is not, since there is in general no independent test of the strength of the derived preference.

This is not to say that there are no preferences, of various degrees of intensity, based on sympathetic identification with others. Of course there are, and they are basic to ethical experience. The point is that understanding, identification, and preference are not related to one another as the World Agent model makes out. Confronted with someone in a dire emergency, I will, if I am a humane person, acquire an overriding preference to help him if I can. That operates through consideration of what it is like for him, a consideration in which some part is played by thoughts of what this or something like it would be like for me. My knowledge of what somebody wants (let us say, that I should help him out of the fire) sets off in me, granted a humane disposition, a desire to help him out of the fire. So there are four relevant truths about me in this situation. First, I know how it is for him and that he wants to be helped. Second, I know that if I were in that situation I should want to be helped. Third, I have a preference now, in my own person, for being helped in such situations. Fourth, being of a humane disposition, I want to help him. On Hare's model, the first of these is equivalent to the second, because being in that situation is taken to mean total immersion. (This brings out clearly how radical the interpretation is. In ordinary life, they are not taken to be equivalent, and much possible comedy lies in their being confused.) On Hare's model, further, I cannot know the second unless the third is true; and, last, “being of a humane disposition” means being disposed to make rational first-personal calculations in which preferences of that transferred kind are given proper weight (as against my own convenience, and so forth).

These connections cannot all be correct. The operations of sympathetic understanding or, as it is often now called, “empathy” have been much discussed in the history of moral philosophy, and various accounts have been given of it. But one thing that must be true is that the insightful understanding of others' feelings possessed by the sympathetic person is possessed in much the same form by the sadistic or cruel person; that is one way in which the cruel are distinguished from the brutal or indifferent. But the cruel person is someone who has no preference to give help (he is not someone who has a preference to give help but finds it outweighed by a preference for enjoying suffering). Yet he certainly knows. Hare indeed says about the connection he makes, that it “is a conceptual truth, in the sense of ‘know’ that moral thinking demands.” But moral thinking demands no sense of “know” except knowledge, and it is a truth, if not a conceptual one, that any knowledge it can use may be turned against it.
I have pursued this question because the operations of sympathy and of the role-reversal test (not necessarily the same thing) are important to ethical thought; and also, more immediately, because they are involved in the influential World Agent interpretation of utilitarianism. This is not the only model for utilitarianism, and other versions of it will escape some of these specific criticisms. But the idea of taking into oneself the world's wants and sufferings and, at an ideal level at least, feeling all of its pains and pleasures as equally close to oneself, is a basic motivation of utilitarianism — the contrast with contractualism is clear in this — and Hare's version of what is in effect the World Agent model brings out with exceptional clarity what is involved.

Utilitarianism is the most ambitious of extant ethical theories. It aims to yield the most definite results and is willing to press them most firmly against everyday ethical beliefs. We must look next at the relation of utilitarianism and of other ethical theories to practice. Why should such theories be granted any authority at all?

**CHAPTER 6**

**Theory and Prejudice**

Ethical theories have to start from somewhere. Earlier I considered ways of their starting outside ethics altogether. I also touched on the idea of starting inside ethics, but merely from the meaning of moral words. I found all of these in varying degrees unpersuasive, and some I rejected altogether. Many would agree with these conclusions, including some writers whose aim it is to construct an ethical theory. They still have to start from somewhere, and the only starting point left is ethical experience itself.

"Ethical experience" can cover many things. There could be a way of doing moral philosophy that started from the ways in which we experience our ethical life. Such a philosophy would reflect on what we believe, feel, take for granted; the ways in which we confront obligations and recognize responsibility; the sentiments of guilt and shame. It would involve a phenomenology of the ethical life. This could be a good philosophy, but it would be unlikely to yield an ethical theory. Ethical theories, with their concern for tests, tend to start from just one aspect of ethical experience, beliefs. The natural understanding of an ethical theory takes it as a structure of propositions, which, like a scientific theory, in part provides a framework for our beliefs, in part criticizes or revises them. So it starts from our beliefs, though it may replace them.

Those initial ethical beliefs are often called in current philosophy intuitions, but that term no longer carries quite the implications it once did. Intuition used to be taken as an intellectual power of
arriving at abstract truths, and its application to ethics lay in the idea that ethical truths could be grasped *a priori* by such a faculty. The philosophers who used this model differed on various questions: what concepts occurred in the truths given by intuition (whether it was to be goodness or obligation, for example); whether those truths were very particular or very general. But in using the notion of intuition, they all supposed that the way in which we grasped those ethical truths was significantly like the way in which we grasp mathematical and other necessary truths. The ethical truths grasped by intuition could provide a starting point for ethical theory, if there was to be ethical theory, but not all believers in intuition in fact wanted ethical theory, since intuition itself was supposed to provide the test or, rather, make tests unnecessary.

This model of intuition in ethics has been demolished by a succession of critics, and the ruins of it that remain above ground are not impressive enough to invite much history of what happened to it. The charges, briefly put, were that it failed to explain how an eternal truth could provide a practical consideration, and that it was wrong in assimilating ethical truths to necessities. If necessary truths such as those of mathematics were seemingly denied by informants from another culture, one would naturally look in the first instance for a better translator, but the situation with ethical beliefs is not at all like that. Above all, the appeal to intuition as a faculty explained nothing. It seemed to say that these truths were known, but there was no way in which they were known. “Intuition” is not much of an explanation when it is applied to what are necessary truths, but with ethical beliefs it is worse, for reasons that once more have to do with cultural disagreement. Little as we know about it, we already know too much about the explanation of ethical beliefs and their cultural differences to accept a model that says there is not going to be any such explanation.

So intuition in ethics, as a faculty, is no more. But intuitions—the beliefs which, when there was supposed to be a faculty, were supposedly given by it—are very much part of the subject. These are spontaneous convictions, moderately reflective but not yet theorized, about the answer to some ethical question, usually hypothetical and couched in general terms. They are often questions about what to do. “What should you do if you could, by switching the points, divert a runaway trolley from one line, where it would certainly kill three old men, to another line on which it would certainly kill one child and a gifted violinist?” This example is not much more fantastic than some that have been offered. But intuitions do not have to be expressed in answers to questions about what to do. Some may be found in our willingness to apply to some imagined situation one of those more substantive ethical concepts, such as those picking out virtues or types of action, that were mentioned in Chapter 1.

There is an analogy that has encouraged the revival of the term “intuition” in these connections. This is its use in linguistics and the philosophy of language to refer to a speaker’s spontaneous grasp of what can and cannot be said in his language, or of what can be correctly said in a particular kind of situation. A competent English speaker has the intuition that it is not correct—that is, it is not English—to say (as I once heard an emigré philosopher of language say), “In English we are not using the present continuous to signify a custom or practice.” Such intuitions are the raw material of a theory of a natural language. We have good reason to believe that it should be possible to form such a theory, giving an account of the rules that have been internalized by the speaker, just because the speaker can unhesitatingly recognize as correct or incorrect in his language sentences he has never heard before. As Noam Chomsky has emphasized, we do this all the time. Moreover, some theorists, notably Chomsky, believe that since any human being can learn as a child any human language, there are grounds for expecting there to be a theory of rules underlying all natural languages, a universal grammar.

How does this linguistic conception of an intuition apply to ethics? There is one kind of intuition relevant to ethics that certainly fits the model, since it is merely an application of it. In the case of the substantive terms for virtues and kinds of action, there is room for linguistic intuitions about the situations they apply to, just because they are general terms in the language with complex conditions of application. (What ethical consequences, if any, follow from people’s capacity to use such terms, differing as they do
between one culture and another, is something we shall come to in Chapters 8 and 9.) With terms of this kind there will be debates about their application at the margin, and these may carry serious practical consequences. They are disputes of the kind familiar in the law, where the issue may be whether a given act constituted theft, for instance. Legal theorists disagree about the exact nature of disputes of that kind and how they are properly decided, so-called legal realists allowing a larger and more explicit role for policy considerations in the decision of hard cases—but all are agreed that there has to be a shared understanding of some core or central cases to make these debates about hard cases possible. To some extent this must be equally so within the less formal structures of the ethical discussions that involve these substantive terms.

In some traditions great weight is laid on this legalistic strain in ethical thought. It is encouraging to objectivist views of ethics, since the core cases are given in an understanding of these ethical terms, and their application to hard cases, though it is a contentious and ethically fraught matter, is constrained by rational criteria of what is and what is not an adequate similarity to the core cases. There can be rational discussion whether a given extension of the term properly bears the spirit or underlying principle of its application to the core cases. Arguments in this style are, in the Catholic tradition, known as arguments of casuistry (the unfriendly use of that term was a deserved reaction to devious uses made of the technique). The trouble with casuistry, if it is seen as the basic process of ethical thought, is not so much its misuse as the obvious fact that the repertory of substantive ethical concepts differs between cultures, changes over time, and is open to criticism. If casuistry, applied to a given local set of concepts, is to be the central process of ethical thought, it needs more explanation. It has to claim that there are preferred ethical categories that are not purely local. They may be said to come from a theory of human nature; in this form, the explanation leads us back to the concerns of Chapter 3. They may be said to be given by divine command or revelation; in this form, if it is not combined with the grounding in human nature, the explanation will not lead us anywhere except into what Spinoza called "the asylum of ignorance." An exponent of the casuistical method could perhaps fall back simply on the idea that the categories we prefer are the ones we have inherited. This has the merit of facing an important truth, but it will not be able to face it in truth unless more is said about ways in which those categories might be criticized.

When we turn away from the use of substantive ethical terms, and merely consider such things as people’s answers to questions about the ethically right thing to do in certain situations, the analogy seems much slighter between, on the one hand, the ability to give "intuitive" (assured and unprompted) answers to these questions and, on the other hand, linguistic competence. The ability to give ethical answers does indeed require some explanation. The presented cases are not exactly like previous cases, and the respondent must have internalized something that enables him or her to respond to the new cases. But it is not obvious what that may be. In particular, it is not obvious that it must be a principle, in the sense of a summary and discursively stateable description that does not rely too much on vague references to degree ("too much," "balances out," "does not pay enough attention to . . ."). In fact there is a dispute in the philosophy of language, to what extent linguistic competence itself, particularly on the semantic side, can be captured in some set of stateable rules. In the ethical case, inasmuch as the problem is seen as the explanatory problem of representing people’s ability to make judgments about new cases, we do not need to suppose that there is some clear discursive rule underlying that capacity. Aristotle supposed that there was no such rule and that a kind of inexplicit judgment was essentially involved, an ability that a group of people similarly brought up would share of seeing certain cases as like certain others.

This is what followers of Wittgenstein are disposed to believe about all human learning. At some eventual level they must be right: understanding a summary discursive rule would itself involve a shared appreciation of similarities. But this conception of the ability to arrive at shared ethical judgments (and the same thing is going to apply to other kinds of practical judgment as well) goes further than that. It is not merely that the ability to use language requires a shared capacity to see similarities, but that the capacity to