ETHICAL THEORY
Ethics is customarily divided into two parts: meta-ethics and normative ethics, with the latter being divided further into normative theory and “applied ethics,” the area with which we are concerned in this volume. This last term may not be especially apt, however, since it suggests a relation to normative theory like that applied to pure mathematics, where theories are derived independently and, only then, applied to cases. When it comes to normative ethics, theories are often formulated and evaluated by reflecting on the ethically relevant features of cases. Thus some philosophers maintain that we can appreciate the general moral relevance of a distinction between killing and letting die (or, more generally yet, between causing evils and letting them happen) by reflecting on a specific case like Judith Thomson’s famous “trolley problem,” in which a driver can choose between letting his runaway train kill a certain number of people and diverting it on to a track where it would kill a smaller number (Thomson, 1976).

In thinking about this case, it seems to be relevant that by diverting the train the driver would be killing people or causing their deaths himself, whereas, if he let the train continue undiverted, he would only be allowing deaths to occur. By seeing this in a specific case, it is argued, we can appreciate a distinction of general theoretical relevance.

Another term for our area, “practical ethics,” avoids these associations but is misleading in a different way, since it suggests that the only cases of interest concern practical questions of what to do. Frequently, what we want to know is not what someone should do (or should have done), but what to think or feel about someone’s character or about his having done something out of certain motives. And there are many other ethical questions that are not primarily practical either, even if they have practical implications: do all living species have intrinsic worth? Is aesthetic appreciation a more valuable form of human experience than the relief of a scratched itch? And so on.
Case Ethics

A better term for our area might be “case ethics.” Just as there is “case law,” the findings of judges about the issues brought before them, including, crucially, the reasoning or ratio that led to their conclusions, so also is there case ethics: our considered judgments about specific ethical issues or cases along with the reasons or principled reflections that underlie our judgments. When it comes to the law, of course, only properly vested judges can render authoritative judgments. With moral and ethical discussion and debate, however, we all have standing to take part, and no individual has the authority to make final judgments, although we may freely accord a kind of authority to those we think especially thoughtful and judicious.

When judges render legal judgments, it is not enough for them simply to find for one side or the other. They must also support their judgments with applicable law and legal principles, aiming to show how anyone reflecting on these laws and principles might reasonably have come to the same conclusion. This rarely involves anything like a deductive proof. As Aristotle remarked, we cannot sensibly demand more logical rigor than “the subject-matter admits of” (Aristotle, 1998). It is enough if judges point to laws, principles, and ideals under which the case might reasonably be subsumed and which, when applied, might be seen to support their finding in the case.

No doubt it is a mistake to view all of ethics on the model of law, as a number of philosophers have noted (Anscombe, 1958; McDowell, 1979; Dancy, 1993). Some ethical questions require a sensitivity and insight that seems more akin to aesthetic appreciation than to anything that can be supported by theoretical principles. None the less, even art critics feel bound to give reasons for their judgments, citing considerations that can help others enter into their reflections and see the work as they see it. Moreover, many of the cases of applied or case ethics with which we are most often occupied are questions of public morality, where we implicitly understand our discussions and inquiries to take place in a democratic society in which everyone has standing to participate. Moreover, when these issues concern moral obligations – the area where, as John Stuart Mill pointed out, we hold one another accountable through formal and informal sanctions – the same considerations that lead us to demand a publicly formulated justification for legal judgments would seem to apply to moral judgments also, if not, perhaps, quite so urgently (Mill, 1957). A restriction of liberty of some kind will seem to be involved, calling for a principled justification that could be seen to be acceptable to our fellow citizens who would, in our view, be bound by them.

Normative Ethical Theory

Philosophers use the term “normative ethical theory” to refer broadly to principles, concepts, and ideals that can be cited in support of ethical judgments about cases. In this broad sense, we commit ourselves implicitly to some theory (or range of theories) whenever we give reasons to support our judgments. Furthermore, there is a sense in which we commit ourselves to the existence of some justifying background theory.
whenever we even make an ethical judgment. This is because of an important feature of ethical concepts and properties that we might call their reason- or warrant-dependence. When, for example, I judge that something is good, I say, not just that I value it, but that there is reason to value it – that valuing it is warranted, an attitude one ought to have. As a logical matter, however, this can be true only if something has other properties: the reasons for valuing it. And such reasons cannot simply consist in the property that it is good, since that is itself the property of there being such reasons. Unlike, say, the property of yellowness, which might attach to something all by itself, as it were, ethical properties require, by their very nature, completion by further properties that are their reasons or grounds. If I judge a certain experience to be valuable, I must think it has aspects that make it good, features that are the grounds of its value. Or if I think that a certain action is morally required, I must think there are certain characteristics of the action and the situation that make it morally obligatory, features that are the grounds of its obligatoriness. And these thoughts commit me to the existence of background normative theories. I am committed to thinking there are truths that relate an experience’s having certain properties to its value, such that any experience that had exactly those (and no other ethically relevant) properties would be valuable also, other things being equal. Or, similarly, I am committed to thinking there exists some valid moral principle that relates an action’s having certain features to its being morally required.

In this broad sense, then, the investigation of normative ethical theories is unavoidable if we are to think about ethical issues with any care. Any judgment we make about an individual case will be no better than the background theories we commit ourselves to in making it. Moreover, there are special considerations that commit us to normative theories of a distinctively robust sort when the judgments we make concern questions of moral right and wrong. This is because, as I noted briefly above, the part of ethics we call morality is modeled on law, even if other parts are not. What is wrong is what we are appropriately held accountable for doing, what warrants blame unless we have some adequate excuse. Practices of accountability are in their nature directive, and frequently even coercive. As with a judge’s legal findings, therefore, we properly feel some pressure to articulate principles (or theories) that are capable both of justifying our judgments and of being publicly addressed to and accepted by other members of the moral community.

Someone we hold accountable for wrongdoing, we think, should be capable, in some sense, of accepting our judgment, of being brought to see that it is a reasonable judgment to have made. This is very different from other ethical assessments, as when, for example, we feel disdain for someone as a coward, or hold some human pursuits to be less worthy than others. Disdain for cowardice does not attempt to direct the coward or to hold him accountable, and because there is no prescriptive or liberty-limiting element, there is no thought that disdain is appropriate only if its object should be able to accept it and see things the same way. To the contrary, disdain may only increase if its object cannot “get it.” If, however, we judge someone to be incapable of assessing his own conduct morally, this can lead us to think that he is not a fit object of moral evaluation, since he is incapable of entering into a mutually accountable moral community. It is therefore reasonable to accept a burden of being able to formulate public justifications for judgments of moral right and wrong that is similar to one we impose on judges’
legal findings. Normative moral theory is in this way part and parcel of public moral discourse.

**Meta-ethics**

Our main interest in this chapter will be with the major normative ethical theories: contractualism, consequentialism, deontology, and virtue theory. Before discussing these, however, we need first to introduce briefly the other main area of ethical theory: meta-ethics. Unlike normative theories, which concern themselves with substantive normative questions – such as “What is valuable?” and “What is morally obligatory?” – meta-ethical theories are concerned with more abstract philosophical issues that underlie these. We can distinguish four different kinds: (a) questions in the philosophy of language concerning the meaning and content of ethical judgments; (b) related issues in the philosophy of mind concerning what mental states ethical judgments express or what it is to hold an ethical view; (c) metaphysical issues concerning the possibility and nature of ethical truth; and (d) epistemological questions concerning the possibility and nature of ethical knowledge and how we can justify our ethical views.

Why, however, should we care about meta-ethics? Some people think that case ethics can be divorced entirely from meta-ethics. They may grant that analyzing ethical cases ultimately calls on normative theory, but hold that this is entirely independent of meta-ethics. I think this view is mistaken and that the sharp separation sometimes made between meta-ethical and normative thought is a distortion both of how the great systematic ethical thinkers (like Aristotle or Kant) proceeded as well as of how we do and should proceed in contemporary moral debate.

Consider, for example, issues that arise in environmental ethics concerning the moral claims that other living species make on us. How much should we weigh harm to other species – either to individuals or to the species themselves – in our moral deliberations? It is impossible to think carefully about such questions without engaging meta-ethical issues. One concerns the nature of harm. To be able to be harmed, something must have a good or welfare. But what is it for something to be good or bad for some being?

On one common view about the good of human beings, a person’s good consists in the satisfaction of desires (alternatively, the desires she would have if fully informed, or would, if fully informed, have for herself as she actually is) (Railton, 1986). This view is sometimes put forward, not just as a normative claim, but as a meta-ethical position concerning what personal welfare or benefit is. But such a meta-ethics of welfare rules out the possibility that any species lacking desire can be benefited or harmed. Roughly, nothing will be good or bad for some being unless it too can be good or bad to it (as it might appear through desire). I believe this meta-ethical theory to be the mistaken result of a line of thought that takes it for granted that a person’s good is what he aims at in so far as he is rational. Once we see that the concept of welfare or benefit has no such privileged status from the perspective of an agent deliberating about what to do, but is one we require, rather, when we care for some being or thing for its sake (ourselves included), we can appreciate why harm and benefit are not restricted to beings with desires (Darwall, 1997, 2002). To have concern or care for members of another species is just to desire for their sake that they do well or flourish. We can
sensibly regard a species as capable of being benefited or harmed, therefore, if we can care for them for their sake.

But, again, what sort of claim does harm make when we understand it in this way? Do we have a moral obligation not to harm members of any species? Or is the fact that an action would harm another person relevant to its being wrong in a way that harm to other species is not? Here again, we cannot answer this question without taking a stand on meta-ethical issues concerning what morality and moral obligation are, if only implicitly. If we understand moral questions broadly enough, then it may seem that harm is harm and is no less morally relevant whether the being harmed is a person or a snail darter. If, however, we think of morality as a system of reciprocity or mutual accountability, where norms of right and wrong mediate a moral community of free and equal moral persons, then harm to other persons will seem to have an intrinsic moral relevance that harm to other species does not. For then what is morally wrong will be what one can be held accountable by others for doing, in accordance with norms that must, in some sense, be acceptable to all from a perspective of equality. So viewed, harm to persons is not simply harm to members of a certain species, but harm to a member of the moral community to whom norms of right and wrong must be justifiable.

This is only one example of how questions of meta-ethics are implicitly involved in issues of normative ethical theory and, therefore, in case ethics. Ultimately, we have no alternative but to pursue philosophical ethics, that is, to attempt to work out a comprehensive outlook that integrates normative ethics and meta-ethics (Darwall, 1998).

**Contractarianism/Contractualism**

We can turn now to a review of different normative theories and begin with one that can be grounded in the meta-ethical theory of morality as reciprocity or mutual accountability just mentioned. This is the idea that whether an action is right or wrong depends on whether it accords with or violates principles that would be the object of an agreement, contract, or choice made under certain conditions by members of the moral community. The general idea can be developed in a variety of ways, depending on how the choice or agreement, the parties who make it, and the conditions under which it is made are characterized. One broad distinction is between contractarianism, under which the choice of moral principles is self-interested, and contractualism, which grounds it in a moral ideal of reciprocity, reasonableness, or fairness.

It may seem strange to think that moral principles can in any respect be agreed upon or chosen. How can a moral proposition be made true by any choice or agreement? Only rarely, however, do contractarians or contractualists claim that right and wrong are determined by actual choices or agreements (Harman, 1975). More frequently, what they hold is that moral principles are those that would be rationally or reasonably chosen or agreed to under certain (frequently, counterfactual) conditions.

**Contractarianism**

Contractarianism was initially formulated by Thomas Hobbes (see Hobbes, 1994). Hobbes begins by considering the situation of an agent deliberating independently of
others from the perspective of his own desires or interests. Each person, he thinks, sees what he desires as good, as giving him reason to seek it. But what results if all of us, together, pursue our respective desires and interests? Although each person’s doing so may actually result in his interests being best promoted, given the conduct of others, it does not follow that everyone’s pursuing his respective interests, rather than everyone’s pursuing some other aim, or acting on some principle other than self-interest, will actually result in everyone’s (or even anyone’s) interests or desires being best promoted. In situations where this is not the case, where the collective pursuit of self-interest leads to an outcome that is worse for each, we have what is known as a collective action problem.

This is illustrated by the game-theoretic example known as the Prisoner’s Dilemma, in which two individuals are jailed on suspicion of robbery. The district attorney tells each that he lacks enough evidence to convict either him or his partner of robbery, but can easily convict each of breaking and entering, giving each a sentence of one year. He offers each a deal: if one confesses and his partner does not, the confessor will go free and the partner will get twenty years. If both confess, both get five years.

Suppose that each cares only about doing the least time. The structure of the situation then is as follows. If A confesses but B does not, then A gets his first-ranked outcome and B his fourth-ranked (worst). And vice versa, if B confesses but A does not: fourth-ranked for A, first-ranked for B. If both confess, both get their third-ranked outcome. And if both do not confess, both get their second-ranked.

What should each do? Reason first from A’s perspective. B will act independently of A and either confess or not. It seems, therefore, that A should confess, since whatever B does, A will do better if he confesses. If B confesses, then A will get his third- as opposed to his fourth-best outcome by confessing. And if B does not confess, then A will get his first- as opposed to his second-best outcome by confessing. So A should confess. A will do better by confessing, whatever B does.

But B’s situation is exactly analogous to A’s, so any reasons for A to confess apply equally to B. So if A would do best to confess, then so would B. It may now be evident why this is called a collective action problem. A’s and B’s actions, although likeliest to achieve the best outcome for each when taken individually, taken together yield an outcome that is worse for each. If both prisoners do what would be best for him, given the actions of the other, both will confess. But that yields each one’s third-ranked outcome, whereas they could have both achieved their second-ranked outcomes by not confessing. Although the jailhouse context makes this sound strange, not confessing is actually the cooperative strategy for A and B. If A and B could cooperate to their mutual advantage, they would both not confess and end up with their second-ranked outcome rather than the third-ranked outcome that independently promoting their interests will achieve.

People cooperate when they forgo the pursuit of their own independent interests and follow rules or roles, the collective following of which promotes everyone’s interests better than would have been done by everyone pursuing their own interests independently. Obviously, cooperation is required for many, many things that are valuable in life, perhaps especially in complex modern societies, in which we cannot assume that genuinely common interests, shaped by common cultural or religious traditions will stretch across all areas of significant interaction. Morality can be thought of as an especially
broad and pervasive form of cooperation. Principles of moral right and wrong would then be whatever rules, specifying requirements, permissions, and so on feature in the broadest possible form of cooperation, namely, cooperation involving not just this or that group, community, or political unit, but all competent human agents. (Actually, Hobbes’s view was that cooperation among large groups was impossible without political authority, since otherwise uncertainty of others’ participation would undermine the assurance necessary for it to make sense for one to do one’s part.)

According to contractarianism, then, whether an action is right or wrong is determined by rules of cooperation of the broadest sort, that is, between all human moral agents. Take, for example, the rule that it is wrong not to come to the aid of others in need, so long as the sacrifice involved is not too great (say, as long as it is not above some level $\alpha$ and/or the ratio of sacrifice to need is not above some level $\beta$). Arguably, there exist some $\alpha$ and $\beta$, such that it would promote everyone’s interests more for everyone to follow the resulting rule, than it would for everyone to pursue their own interests independently. If that is so, then, for starters, contractarianism will hold that it would be wrong not to follow this rule.

To a first approximation, contractarianism holds that what it is right to do depends on what rules it would be in everyone’s interest for everyone to accept and be guided by in their deliberations and moral practice. However, what if various different possible rules for a given kind of situation have the property that everyone’s interests would be promoted better by everyone’s following that rule than they would be if everyone attempted to promote their own interests independently? This is where the idea of an agreement or contract enters the contractarian picture. Taking as a benchmark the “no agreement” point in which all regard themselves as bound by nothing but their own interests and values, contractarians treat the question of which principles we actually are morally bound by as the solution to a rational bargaining problem from this benchmark, in which we all have a greater interest in agreeing to some mutually advantageous principles, thereby avoiding the “no agreement” point, but have differing interests in exactly which principles are actually agreed (Gauthier, 1986). How favorably the resultant principles treat the different negotiating agents will depend on who has the most to lose if there is no agreement. Consider, for example, what principle of mutual aid would be agreed to. If those with fewer resources and greater vulnerabilities have more to lose from the lack of agreement than those with greater resources and fewer vulnerabilities, then rational bargaining may lead to a less onerous principle of mutual aid than would result if everyone were as vulnerable as those with less.

This, then, is contractarianism’s basic framework for assessing moral issues. To work out our moral obligations in a specific case, say, the obligations of rich and poor countries in reducing global warming, we have to think about what agreement on principles for dealing with the issue would result from a negotiation from the “no agreement” point in which each party attempts to advance its own interests and values. In acting on the principles that would be agreed, however, the parties are not simply promoting their interests; they are cooperating. Cooperation promotes everyone’s advantage, but, as in the Prisoner’s Dilemma, it does so by requiring individuals to forgo promoting their own interests. Each would prefer schemes in which the necessary sacrifices are borne in greater measure by others, but each party is prepared to do its part as required by principles of cooperation that everyone could rationally agree to.
Contractualism

Contractualism has a similar structure. It too understands principles of right conduct as the object of a rational agreement. But whereas contractarianism takes moral principles to result from rational bargaining, contractualism sees the agreement on principles as governed by a moral ideal of equal respect, one that would be inconsistent with bargaining over fundamental terms of association. From contractualism’s point of view, the problem with contractarianism is that it must assume that individuals have, in effect, a moral claim to the resources they would have if there were no agreed rules of cooperation. Otherwise, the rules that result by bargaining from that position will have no moral force. But why assume that people have such a moral claim? From a moral point of view this seems entirely arbitrary, unless some background theory of natural rights is assumed. And contractarianism cannot justify that assumption, since its own moral force would have already to depend upon it.

This problem emerges from another direction if we consider how individuals might get from self-interested practical reasoning to contractarian moral reasoning. If each agent reasons in terms of her own independent interests, then, from a situation in which there are no established rules of cooperation, she will think it rational to bargain to an agreement with others to be bound by certain rules. But how can this give her a reason actually to follow the rules? The reasons of interest she has for agreeing to follow the rules cannot give her a reason actually to follow them since the whole point and function of rules of cooperation requires that they constrain her pursuit of her interests. For her to be able to reason as these rules require, she must already accept moral reasons of cooperation. It may even be that it is in her interest to be someone who does accept contractarian moral reasons of cooperation, but while this would give her reasons to want to accept the moral reasons, she could not accept the moral reasons for these reasons.

The animating idea of contractualism is implicit in Kant’s “kingdom of ends” formulation of his Categorical Imperative. Kant maintains that anyone subject to the moral law must be able to be regarded also as “making the law” (Kant, 1998). Only thus can the moral law be thought of as a common law for a community of free moral agents, subject only to laws they legislate themselves. This is a version of Rousseau’s idea of legitimate political community as an association in which each, “while uniting with all, nevertheless obeys only himself” (Rousseau, 1987). According to Rousseau, this is only possible if laws express what he calls the “general will,” the will of each as a free and equal member. Similarly, Kant conceives of moral laws as “made” by each moral agent when each would “legislate” it as a free and equal member of the “kingdom of ends.” Here we have the central difference with contractarianism. Moral principles of right are not rules that individuals would prescribe, and attempt to gain acceptance for, from their different individual perspectives, bargaining out of self-interest. They are, rather, rules individuals would prescribe (and agree to) from a common perspective as one free and equal person among others.

But how, more specifically, is this perspective to be understood? Kant gives some hints, saying that we can conceive a “systematic union of rational beings under common objective laws,” only if we “abstract from the personal differences between rational beings, and also from all the content of their private ends” (Kant, 1998). This suggests the contemporary contractualist (John Rawls), idea that principles of justice
are those it would be rational to choose in an “original position” behind a “veil of ignorance” regarding any features that individuate different persons or their societies (Rawls, 1971, pp. 136–142). In particular, the choosing parties are ignorant of their individual resources, abilities, talents, gender, race, socioeconomic position, and their own interests or individual values. Rawls assumes that the parties have an interest in autonomously choosing and pursuing their interests (whatever these turn out to be) and, therefore, that they value the “primary goods” that are necessary for these: freedom, opportunities, wealth, and the “social bases” of self-respect.

Rawls’s idea then is that justice is determined by whichever principles the parties would choose from behind the veil of ignorance, that is, as one free and equal person among others. Rawls does assume that this choice is self-interested within the constraints placed by the veil, but this does not reduce moral reasoning to self-interest in any way. To see this, suppose the parties were motivated not by self-interest but by concern for a single other individual. Since the veil of ignorance deprives them of any information that would let them tailor principles to any particular person’s interest, there is no functional difference between assuming the parties to be self-interested and assuming them to be trustees for another individual. The original position is, in effect, the perspective of an arbitrary, free and equal individual.

Rawls argues that the rational choice in the original position would be two principles of justice, ranked in order of priority: a first that requires the existence of certain basic civil and political rights and freedoms and a second, the “difference principle,” which assures fair equality of opportunity and that remaining primary goods, such as wealth, are distributed by the basic institutions of society in ways that work to the greatest advantage of those who are least advantaged. In effect, this says that inequalities are justifiable only to the extent that they work as a social resource (for example, by providing incentives) from which everyone benefits, including the least advantaged. If we take seriously the possibility that we could be anyone, and have no way of estimating probabilities of ending up in any particular position, the rational thing would be to protect against the worst possibilities by choosing the two principles, including the difference principle.

Rawls put these ideas forward as a theory of justice: “justice as fairness,” he called it. More recently, he has stressed that it is to be understood as a political, rather than a more general moral, theory (Rawls, 1993). In his earlier work, however, he suggested that it might also be conceived as a moral theory: “rightness as fairness” (Rawls, 1971, p. 111). To do this, we would have to ask: what principles of individual conduct (or of the conduct of groups, nation states, and so on) would it be rational to choose to govern everyone’s conduct (as well as moral criticism and practices of accountability concerning it), from the original position. This would then give us a framework within which to consider which principles ought to apply to any specific case we might have under discussion.

A second contractualist approach can be motivated by thinking about what it is to make a claim on someone as an equal. When one person claims something in this way, she attempts to give another person a reason based on her needs as an equal. It is as if she says, “This is a reasonable claim for you to grant to me, as you can see were you to put yourself in my shoes and consider that it would be reasonable for you to make it of me.” Such a claim implicitly invokes the idea of principles of conduct that reciprocally recognizing equals can accept or, at least, not reasonably reject. In developing such
a contractualist approach, T.M. Scanlon assumes a community whose members wish to be able to justify their conduct to each other by principles that others could not reasonably reject, in so far as they also have this aim (Scanlon, 1998, pp. 147–257). Principles of moral right and wrong can then be thought of as norms that structure a mutually accountable community of equals.

To apply this criterion, we must make judgments about what is reasonable. How can we make these? There seems no alternative to putting ourselves into others’ shoes and seeing whether we would regard a certain claim, or objection against a proposed principle, as one we would reasonably make if we were in their situation. This is a complex judgment. It is not simply the prediction that we would make the same claim or objection. We might think we would, but that it would be unreasonable. To make the requisite judgment, it seems, we must attempt to enter into the other’s perspective impartially, as anyone, to see whether we would endorse the claim or objection as a reasonable one to make to another, reciprocally recognizing equal. Suppose, for example, we are trying to determine what principles should govern reduction in nations’ use of fossil fuels to combat global warming. Contractualism will hold that developed and developing countries alike should govern their conduct by principles that none of them could reasonably reject. If a proposed standard is in fact rejected by developing countries, for instance, it is necessary then to judge whether this would be a reasonable objection to make were one in their shoes. And vice versa for standards that might actually be rejected by developed countries.

**Consequentialism**

Whereas contractarianism and contractualism begin from within conceptions of morality, as mutually advantageous cooperation or as reciprocity between equals, respectively, consequentialism begins with values it holds to be prior to morality. Even if there were no moral right and wrong, some things would still be good and others bad. When we judge the pain and suffering caused by a cataclysmic earthquake to be bad, for example, we are not making a moral evaluation, even if our judgment has implications for morality. Neither the pain nor the earthquake need have involved agency or character in any way. Rather, we are judging that it is a bad thing that the suffering happened, that such suffering is a bad state of affairs, a bad thing to happen. The idea is not just that suffering is bad for the sufferer, but that it is a bad thing to occur period. As these values and disvalues are independent of morality, they are called non-moral.

Consequentialist moral theories start with a non-moral value theory: a normative theory of which states of the world (things that can happen) have intrinsic value, which have disvalue, and some account of how these values compare, either with an ordinal ranking or with some cardinal metric. What makes these values non-moral, again, is that they are not evaluations of moral agency or character, but of outcomes or states – ways the world might be. Of course, such states might include agency and character. But even here the evaluation of the state (as something that happens) can be distinguished from the evaluation of the act or character trait that is a constituent of that state. Thus one might consistently think it would have been good if Hitler had been assassinated, that that would have been a good thing to have happened, say because of
the lives it would have saved, even if such an assassination would, perhaps, have been morally wrong. As we will see, consequentialists might deny that such a killing would be wrong, but the point is that there would be no incoherence in holding it to be wrong and, at the same time, thinking that the state of the world of Hitler’s being assassinated would, on balance, have been a good thing to have occurred.

Consequentialist moral theories all agree that the moral rightness and wrongness of acts are determined by the non-moral goodness of relevant consequences. There are, however, two kinds of issues on which consequentialist theories divide. First, and most obviously, they can disagree by being based on different theories of non-moral value. A consequentialist with a hedonist value theory, according to which pleasure is the only intrinsic good, will disagree, for example, with one who holds that preserving species or, perhaps, historical cultural treasures can be good things in themselves. Second, consequentialist theories can also disagree by holding that consequences of different sorts are relevant to determining moral right and wrong. Act-consequentialism holds that whether a given act is right depends on the value of the consequences of that act, compared with the value of the consequences of any other act the agent could do in the circumstances. According to rule-consequentialism, on the other hand, the rightness of acts depends on the consequences, not of the act, but of the social acceptance of a rule requiring, forbidding, or permitting the act, compared with the consequences of accepting other possible rules for that kind of case. If accepting a rule requiring an act of that kind would have the best consequences, then the act is morally required. And consequentialism can take other forms too.

All forms of consequentialism, however, understand moral evaluation to be an assessment of instrumental or extrinsic value at the most fundamental level. All are based on theories of the intrinsic, non-moral value of outcomes, and all assess the moral status of acts and character by determining which acts, social rules, or traits of character are the best instruments for promoting the most valuable states. For act-consequentialism, a morally right act is the agent’s best available instrument for producing the most valuable states. And rule-consequentialism judges the rightness of acts by the verdicts of socially realizable rules that are, via their participation in social practices of moral reasoning and criticism, the best instruments of that kind for producing non-moral value.

In principle, virtually any theory of outcome value can be harnessed to a consequentialist moral theory. Historically, however, consequentialism has been advanced most frequently by philosophers who have thought that valuable outcomes must somehow involve the lives of conscious beings. We might call benefit consequentialism the view that valuable states all concern the good or welfare of some being or other, and that moral assessment must ultimately be based on this. It is possible to believe, however, that something can benefit or harm a being by affecting something other than the quality of its experience or conscious mental states. For example, perfectionists sometimes assert that a being’s approximating an ideal for its kind is intrinsically beneficial to it. This is what leads to the conclusion of Aristotle’s famous “function” argument that human good or flourishing consists in excellent, distinctively human activity (Aristotle, 1998). By and large, however, benefit consequentialists have tended to hold that people are benefited or harmed, respectively, by what positively or negatively affects their mental lives, that is, to hold either hedonistic or desire-based forms of consequentialism.
The most popular form historically has been utilitarianism, which is distinguished by three features. First, utilitarians are benefit consequentialists who, because they hold either hedonistic or desire-based conceptions of benefit, maintain either hedonistic or desire-based consequentialism. Second, utilitarians hold that the non-moral value of outcomes is determined by summing the benefits and costs to all affected parties. And, third, utilitarians believe that the moral rightness of action or the moral goodness of character traits depends on what would produce the greatest overall value, determined by such a sum. The classical hedonistic utilitarian formulation, in Bentham for example, holds that happiness is an experienced state and that people can be benefited only by the intrinsic qualities of their conscious lives; that is, by the degree of pleasure that they experience compared to their pain or suffering (Bentham, 1970). A different kind of utilitarianism, based as much, perhaps, on a free-standing value of autonomy as on a conception of happiness, holds that an individual’s welfare is determined by his own desires and preferences. Since people can have preferences for things other than the intrinsic qualities of their own conscious states, this desire-satisfaction form of utilitarianism has rather different implications from a hedonist version. For example, someone might strongly desire the survival of a certain wilderness area. A desire satisfaction form of utilitarianism would weigh this fact in favor of saving the area even if saving it made no contribution to the quality of any being’s experience (say, because the individual in question did not know anything about the area’s survival).

Although consequentialists have usually been utilitarians or benefit consequentialists of some sort, there is nothing in the logic of consequentialism that restricts it to these versions. Philosophers have frequently argued that such things as knowledge, understanding, friendship, love, beauty, and artistic and other cultural activity and creation have intrinsic values that cannot be reduced to the benefits they bring to human (or other sentient) life. After all, some of our deepest satisfactions seem themselves to involve the appreciation of these values, so the values cannot wholly consist in these satisfactions. When it comes to controversial issues of case ethics, for example in environmental or medical ethics, it is open to consequentialists to argue that relevant values include such things as the existence of a species or of a relationship of a certain kind between doctors and patients. The consequentialist’s test for moral relevance will simply be whether a given state of affairs’ existing makes a positive contribution to the value instantiated in the world, whether it is a good or bad thing that it exist or happen.

This structure enables consequentialism to take account, in principle, of a wider range of considerations than contractarianism or contractualism. Most notably, there is nothing in the consequentialist conception of morality that ties it specially to the condition of other members of the moral community, or that restricts its consideration to human beings in any way. If pain or suffering is a bad thing, then, it seems, it would be bad whether the being that suffers is capable of moral agency or not. For this reason, advocates for the interests of animals frequently cite Bentham’s dictum: “The question is not, can they reason? . . . but, can they suffer?” (Bentham, 1970, p. 282). And consequentialists who believe that intrinsically valuable states are not restricted to those in which beings benefit can take advantage of a wider array of considerations yet. Of course, moral reasoning is not simply a matter of deploying rhetorical resources. Any moral conception will face the burden of defending the relevance of the considerations it advances and, in the end, anyone who adopts that conception will need to think
through how the moral relevance of such considerations can be situated in a philosophically adequate conception of moral obligation.

An important feature of any consequentialist view is what philosophers call the agent-neutrality of its fundamental values. Since the values are held to derive simply from the existence of the relevant states, they provide a justification for any actions, policies, or practices that might promote them, irrespective of the agent’s relation to these states. An example will clarify this idea. Suppose you think that among the intrinsically bad things that can happen is someone’s being betrayed by a friend. The thought here is not that it is wrong to betray friends, or even that this is a bad thing to do, but that someone’s being betrayed by a friend is a bad thing to happen. And again, it is not just that this is bad for the person being betrayed, but that it is a bad thing period for someone to be hurt in this way. If you think this is a bad thing to happen, you should also think that it would be good, other things being equal, for actions, policies, and practices to be taken that would prevent this. Suppose that there are two people, Jones and Smith, who are contemplating betraying their friends. Suppose also that circumstances are such that if you betray your friend, Jones and Smith will be so horrified that they will not betray their friends, although they would have otherwise. We can now put the idea of agent-neutrality this way. From the point of view of the intrinsic badness of friends being betrayed, you would seem to have reason to betray your friend since it would go further towards minimizing the intrinsically disvaluable states of betrayals than would your not betraying your friend. Of course, this apparently runs against moral common sense. That is because we commonly believe that friends have duties to each other that are agent-relative rather than agent-neutral, that a moral agent has a duty not to betray his friends that is not reducible to preventing the (agent-neutral) evil of friends being betrayed.

The idea that there are moral obligations that are agent-relative in this way is a hallmark of deontological moral theories. According to deontologists, agency and action are not simply instruments for producing valuable states. Rather, actions are based on reasons and principles, and some important moral principles crucially involve the agent’s relation to various persons (or other beings) in the outcomes she affects. It is commonly thought to be a wrong-making feature of an action, for example, that it will involve one’s (that is, the agent’s) harming others, betraying a friend, breaking a promise, divulging a confidence, and so on.

Some consequentialist theories, although not all, can agree with these aspects of moral common sense. Rule-consequentialism will agree if, and only if, it produces the greatest overall value (assessed agent-neutrally) for there to exist social practices of moral criticism and psychological patterns of moral reasoning that are themselves guided by agent-relative rules according to which it is a wrong-making feature of an action that it involves one’s (the agent’s) harming others, betraying friends, and so on. Moreover, it is widely agreed among consequentialists that this is so. Consequentialists generally agree that the most effective way to produce the greatest overall value is indirectly. Were everyone to be guided by act-consequentialism in their deliberations and moral criticism, the results would be much worse for many different reasons. Shared rules are necessary to coordinate complex cooperation, establish reliable expectations, diminish self-serving rationalizing and special pleading when the long-run effects of particular actions are unclear, and so on. In the end, however, even rule-consequentialists will agree that the fundamental reason for accepting such
agent-relative rules and principles is that this is instrumentally useful in promoting states whose value is agent-neutral.

**Deontology**

Deontological theories depart from consequentialism on this fundamental point. They hold that what is morally right and wrong is not determined at any level of analysis by what would promote the best outcomes or states, assessed agent-neutrally. They may even be skeptical of the very possibility of pre- or non-moral evaluations of states that are both agent-neutral and morally relevant. Deontology disagrees with act-consequentialism in holding that producing good or bad outcomes is not the only thing that tends to make an act right or wrong. Deontology also disagrees with rule-consequentialism in holding that the reason why this is so is not that believing it to be so itself produces the best outcomes. Deontologists hold that at least some fundamental moral principles or ideas are agent-relative “all the way down.”

Contractualism is one example of a deontological theory, since it holds that moral principles are grounded in the fundamental, agent-relative idea of living with others on terms of mutual respect. As this shows, our categories are overlapping in various ways. The situation is roughly as follows. Deontology and consequentialism, as we are defining them, are mutually exclusive and exhaustive of normative theories of moral right and wrong. Generally, contractualism and contractarianism are deontological theories, for reasons explained in this paragraph. Virtue theories are a mixed bag. Some are advanced not as moral theories at all, but as supplements or, in some cases, as replacements when put forward as part of a critique of morality (see section on “Virtue Theory”). Some are deontological theories. And some are consequentialist in at least some important respects. Thus Francis Hutcheson (see below) advanced a virtue theory according to which universal benevolence was the highest virtue and argued for act-utilitarianism on the grounds that this theory comports with the deliberative reasoning involved in this highest virtue. But deontological theories and principles are often defended directly, without attempting to ground them in some theory or idea that is held to be somehow more basic. Historically, these versions of deontology have been called intuitionist or species of intuitionism. What characterizes intuitionism in general is the view that there is an irreducible plurality of different right- or wrong-making features whose moral relevance cannot be derived from some more fundamental principle or reasoning but can only be confirmed by moral reflection or “intuition.” This might be done directly as when, for example, it can seem obvious on reflection that the fact that an action would amount to a betrayal or a broken promise must count against it morally. Or it can occur in thinking about or analyzing a specific case as, for instance, when we reflect on the “trolley problem,” it can seem evident that causing harm and allowing it to happen are morally different.

Another example, defended by some deontologists, is the “doctrine of double effect,” according to which there is a moral difference between causing harm or evil as an unintended side-effect of an intended action or policy and intending the harm or evil directly, either as an end or as a means to an end. Thus, although it is a terrible thing whenever innocent civilians die during wartime, for example, when bombing military targets causes even a small number of casualties, it would seem to be worse to try to
kill the same number of civilians directly, even if doing so would produce the same valuable end of victory over a repressive, aggressive regime. One issue of case ethics in which this principle has played an important role is the controversial issue of abortion. Since abortion aims directly at the death of the fetus, it is sometimes argued that it is morally worse than another action would be which caused the fetus’s death only as an unintended side-effect. While it might be permissible to perform a medical procedure that is necessary to save a pregnant woman’s life even at the risk of killing the fetus, it is argued that aborting a fetus to save the woman’s life is morally wrong none the less because it is an impermissible intentional killing.

Deontological intuitionists have defended a wide variety of independent principles or doctrines of right- or wrong-making features of conduct. In addition to the doctrine of double effect and the distinction between “doing and allowing.” there have been claimed to be: duties of beneficence or mutual aid, duties of non-maleficence (“do no harm” – along with the idea that these are weightier, other things being equal, than duties of beneficence), duties of gratitude for benevolence shown, duties of restitution for wrongs and injuries done, duties of fidelity relating to promise and contract, duties of personal relationship (including those of friends, parents, children, family members, and caretakers more generally), professional duties, duties owing to desert (what people deserve), duties of reciprocity and fair play, further duties of justice, duties to other animals (to the extent that these have not been included already), duties to ourselves, and various others. Of course, intuitionists do not agree about every doctrine or principle, not even about all we have mentioned. But they are none the less agreed that some such list of independent principles or doctrines is correct and that the principles on the list cannot be derived from some more fundamental principle or theory, such as contractualism or, even more so, consequentialism.

Both intuitionist and contractualist deontologists hold that the right is, in Rawls’s phrase, “prior to the good” (Rawls, 1971, pp. 30–33). They believe that any attempt to derive the right from agent-neutral outcome value is bound to fail, since the question of what it is right or wrong to do is one that faces agents from a place within the world, defined by a complex set of relations to others who make widely varying claims on us owing to these different relations. What states of affairs would be good to exist, considered as from some agent-neutral observer’s standpoint, may, if there are such values, be among the considerations that are relevant to what a moral agent should do. But our duties depend on our place within the states an observer might contemplate, specifically, on the myriad relations we stand in, and that our actions bring us to stand in, to other agents and patients, to our own past acts, to the histories of those with whom we interact, and so on.

Many of these relations were listed in passing above, but we should have them before us more explicitly. We cannot begin to exhaust them, but it will be helpful to give some idea of their range.

1 Duties of beneficence and non-maleficence. Like consequentialists, many deontologists believe that how our actions affect the good of others (other persons, at least, and perhaps any other being who can have a welfare or good) always has some relevance to what we should do morally. But the relevance, again, is not just that these are valuable or disvaluable outcomes we can promote or prevent. It also matters what antecedent relations we have to the affected parties (and
what relations our actions bring us to have). Harming another is worse (an injury done to the other), other things being equal, than forbearing to benefit. It is not just the causing of a disvaluable state; it is doing harm to some being. Doing harm is worse, also, than failing to prevent it. And directly intending harm is worse than causing it as an unintended side-effect.

2 **Duties of special care.** Various special relations of caretaking give rise to special obligations of beneficence. Thus parents have obligations to promote the welfare of their children that are much greater than the duties of beneficence we have to others in general. And similarly for trustees and other relations of more specialized concern, such as doctors, teachers, and so on, who are responsible for their patients’ or students’ medical or educational welfare.

3 **Duties of honesty and fidelity.** Obligations not to lie or intentionally mislead, to keep promises, not to violate contracts, and, more generally, not to encourage expectations we intend not to meet all fit under the general category of keeping faith and not violating trust. Various personal relationships, like those of friends, lovers, and spouses, can be placed under this rubric as well.

4 **Duties deriving from agents’ and patients’ histories of conduct.** When we wrong and injure others, we incur duties to them to acknowledge fault and offer restitution (*agent-fault*). When others benefit us, we acquire duties of gratitude toward them (*patient-benevolence*). A person’s past conduct may call for some appropriate response, especially from those who have special responsibilities to respond appropriately to merit and desert, such as judges of various kinds (*patient desert*).

5 **Duties of reciprocity and fair play.** There is a duty to do one’s part in mutually advantageous cooperation, especially when one voluntarily accepts cooperative benefits. Contractarians/contractualists see this duty as fundamental. For intuitionists, it is simply one independently important duty among others.

6 **Further duties of justice.** Various further duties of justice derive from political relations; for example, from that of equal citizenship. Here we have duties to support a just political order that establishes and protects basic rights and achieves distributive justice. Where actual political relations are lacking, as in the international context, justice may require that we do our part to help establish justice more widely through more extensive political forms.

7 **Duties to other species.** Here again, our duties depend on complex relations. In addition to duties of beneficence and non-maleficence, we can acquire special obligations to members of certain species owing to our history of interaction with them. Even if other animals cannot be full partners in cooperative schemes, we can acquire duties to them owing to the ways in which we have involved them in our lives and ourselves in theirs. Pets are an obvious example, but no less significant may be cases where species are themselves shaped and cultivated for human purposes in ways that give them special needs and vulnerabilities.

This list, again, is hardly exhaustive. It should be obvious even at this point, however, that the decisions we face in actual cases will inevitably involve, not just a single principle or right- or wrong-making consideration, but complex combinations of principles or considerations. Since intuitionistic deontology rejects the idea that some overarching principle, idea, or process of reasoning exists in terms of which these principles or
considerations might be integrated or prioritized, how do intuitionists believe that the messy business of moral reflection is to proceed in thinking about concrete cases?

W.D. Ross distinguished between the claim that a given duty or right- or wrong-making consideration holds prima facie, that is, other things being equal, and the proposition, that, in some actual circumstance, something or other is our moral duty, all things considered, or, as he put it, sans phrase (Ross, 1963). (Since “prima facie” suggests something epistemological, philosophers nowadays are as likely to use the term “pro tanto” [“as far as it goes”]. The central idea is that a right- or wrong-making consideration is one that makes an act right or wrong, other things being equal, such that were that the only morally relevant feature then the action would be right or wrong, all things considered or “sans phrase.”) It was claims of the former sort that Ross held to be self-evident to intuition. To render a moral judgment in any actual case, however, it is necessary to reflect on all of the morally relevant features and, moreover, on how they interact. To take a familiar kind of case, one may have promised to do something of relatively mundane importance only to find oneself placed in a position to give another aid without which he or she may die. Here both promise and need continue to have weight, but one is weightier and so overrides. It would be wrong not to render life-saving assistance, but the moral force of the promise continues, giving rise to a residual obligation to compensate the promisee in some way. But this is not the only way in which moral considerations can interact. Sometimes one consideration can wholly defeat another. When, for example, a benefit one is in a position to provide is tainted by injustice, this may cancel the positive reason to provide the benefit and not just outweigh it.

Ultimately, on an intuitionist picture, there is simply no substitute to carefully considering ethical cases in all of their complexity. Analyzing or “factoring” a case into various right- or wrong-making features or prima facie duties is an important part of the process. But even here, because these can interact in ways that intuitionists believe defy general formulation, one can do no better than to come to grips with these complex interactions in a way that leads to a reflective sense of what moral verdict they will ultimately support.

**Virtue Theory**

Contractarianism/contractualism, consequentialism, and deontology are all moral theories. Moreover, since case ethics is predominantly concerned with practical questions, we have been considering these theories as accounts of morally right conduct. An approach called virtue ethics, frequently associated with Aristotle, is orthogonal to these theories in both respects. First, virtue is concerned primarily with character rather than conduct – with how we should be rather than what we should do. And, second, virtue ethics is frequently advanced, not as moral theory, but as accounts of other, ethically deep aspects of human life that are, it is sometimes argued, potential rivals to and perhaps replacements for morality and its distinctive forms. The conception of morality as a set of universal and finally authoritative norms or laws by which all moral agents are categorically obligated is far from the only form that ethical reflection can take (Anscombe, 1958; MacIntyre, 1981; Williams, 1985; Slote, 1992). The modern idea of morality derives from a distinctive historical tradition, the

Theories of Ethics  29
Judeo-Christian-Islamic idea of divinely ordained law, to which it is a secular successor. Some philosophers have argued that this conception of morality is seriously defective in various ways and that our ethical reflections would more profitably take other forms. Several who have made this argument have looked to Aristotle’s *Nicomachean Ethics* for a more promising ethical conception. For Aristotle, the fundamental question is not, as for Mill, Hobbes, or Kant, “What is the fundamental principle of moral right or duty and how might this be defended philosophically?” Aristotle asks, rather, “What is the goal of human life? What kind of life is best for human beings?”

Aristotle’s is a distinctive kind (a paradigm, perhaps) of non-moral virtue ethics – “non-moral,” again, because, although Aristotle’s translators frequently use “moral virtue” to signal that he is talking about excellences of character that are concerned with choice, Aristotle does not relate these virtues to any conception of a moral law under which all are accountable as equals. Virtues, for Aristotle, are dispositions to choose what is fine or noble (*kalon*) for its own sake, and to avoid what is base. The operative notion is what Nietzsche called a “rank-ordering” ideal with respect to which one can be better or worse, not a norm or law that one complies with or violates. For Aristotle, the operative ethical emotions are shame, esteem, pride, and disdain or contempt, not guilt, respect, self-respect, and moral indignation.

Virtues are excellences (*areté*), traits, that is, that make something an excellent instance of its kind. It is a virtue in a knife, for example, that it have a sharp edge so that it can cut well. In general, we reckon which traits are excellences (excellent-making) in relation to a thing’s function (*ergon*) or characteristic activity. As Aristotle believes that the characteristic activity distinctive of human beings is action (*praxis*) that expresses a distinctively human form of choice (of actions, valued in themselves as noble or fine (*kalon*)), he concludes that the virtues are traits of character, that is, settled dispositions to choose certain actions and avoid others as intrinsically noble or base. We might put his point by saying that human excellences are states of character concerned with choices that are themselves guided by an ideal of human excellence.

In general, a non-moral virtue ethics is any such (non-moral) human ideal. Although it may be tied, as is Aristotle’s, to a teleological or perfectionist view of human nature, according to which there is something that human beings are inherently for or to be, it need not be so based. A non-moral virtue ethics may be put forward simply as a normative view about which traits in human beings are worthy of esteem (or disdain). Analogously, a moral virtue ethics is a theory of what is worthy of distinctively moral esteem, that is, traits that are worthy of esteem *in a moral agent*. Exemplars of such a view can be found in Leibniz and the eighteenth-century Scottish philosopher, Francis Hutcheson. Thus Hutcheson argued that the basic moral phenomenon is esteem for benevolence, the desire to benefit others and make them happy. Moral esteem, he held, is not primarily for any outcome, but for a motive or trait of character, namely, the desire to produce good outcomes for human beings and other sentient beings.

There are various ways in which a theory of virtue ethics might bear on questions of case or practical ethics. First, non-moral virtue ethics reminds us that questions of right and wrong are far from the only, or perhaps even the most important, ethical questions we can ask in specific cases. Thus, it might be that failing to provide significantly more aid to relieve world hunger and suffering, although not seriously morally wrong, none the less manifests vices of complacency and self-satisfaction. Or even if, suppose, the environment is not something that can be wronged or unjustly treated,
clear-cutting may still manifest an inappropriate attitude toward the environment or unlovely traits that are at odds with living a fully satisfying human life. Second, conceptions of the virtues can give us an independent purchase on what action it is appropriate to take in specific cases. In considering what to do, it may be helpful to ask what a virtuous person, or someone with a specific virtue (say, generosity), would do in that case. This may simply be a useful heuristic, but it may also reflect the Aristotelian view that there is no way of formulating ethical insight in terms that can be grasped and more generally applied by someone who lacks the wisdom or “sense” of the virtuous person. As Louis Armstrong is reputed to have said about jazz, “If you have to ask, you’ll never know.”

Third, virtue ethicists may put forward conceptions of virtue, not simply as guides to appropriate (or morally right) action, but as accounts of what makes an action appropriate or morally right. Thus, it can be held that an action is the right or appropriate thing to do in some case or circumstance just in case it is what the virtuous person would (characteristically) do in that circumstance (Hursthouse, 1999). Such a view might depart from the letter of Aristotle’s position, since he identified virtues as settled dispositions to choose specific actions for their own sake (as noble). This would seem to make which traits are virtuous depend on which actions are noble, not vice versa. Nevertheless, since it would hold that no access to the appropriateness of action is possible save through the wisdom or conduct of a virtuous person, such a view would remain quite close to Aristotle’s in fundamental spirit. What is common to any virtue ethics is the idea that guidance on controversial questions of case ethics can be gained only by looking to the virtues or the virtuous person as a model.

Writers on case ethics therefore look to virtue ethics less frequently than they do to other ethical theories, especially when they are concerned with issues of moral right and wrong. If judgments of moral obligation are implicitly directive, holding others accountable for compliance, it will be reasonable to demand that justifications for these judgments be formulated in terms that those subject to the judgments can, in principle, accept. It is this demand that has led to the kinds of normative moral theories that have been advanced by contractarian/contractualists, consequentialists, and deontologists. In each case, the goal has been to articulate action-guiding principles of right conduct that can be grasped and applied without any special virtue other than the judgment of normally competent moral agents.

References

Further Reading