CHAPTER IV

"Formalism"

The Distinction Between Formalism and Teleology

Traditionally, the most fundamental division of theories on the "normative" question of ethics is doubtless that between "teleology" and "formalism" or "deontology." According to this division of the subject, teleological theories are those according to which there is a single criterion for making moral assessments of acts, namely, their productiveness of good or evil consequences, while "formalist" theories hold that something besides the consequences is also relevant. This "something-else" is the "intrinsic nature" of the acts in question.

In this section, I shall contend that this distinction is misleading and unsatisfactory. We have not been told much about the meaning of the phrase, "the intrinsic nature of the act," other than that it is something besides the "consequences"; unfortunately, the term 'consequences' seems to be in some important way relative to the term 'intrinsic nature.' A close inspection of these expressions will reveal that the dichotomy fails to divide theories in the ways intended. The application of utilitarianism, which is supposed to be (and surely is) a "teleological" theory, requires a distinction between acts which are right or wrong on account of their "intrinsic natures" and those which are right or wrong on account of their "consequences," as indeed would any theory.

1 A longer discussion of this subject is to be found in my article, "Utilitarianism and Formalism," Australasian Journal of Philosophy (May, 1965).

2 We will confine the term 'formalism' to the present matter, reserving 'deontology' for the subject of duties and rights, which will be discussed later.
MORALITY AND UTILITY

Let us, then, take a look at this term, "intrinsic nature." Clearly, it is used in rather the way in which "essence" is used. Just as the "essence" of anything is what makes it what it is, so the "intrinsic nature" of an act is what makes it the kind of act it is. Now, what is the sense of the word 'makes' here? Usually, it is to be understood in a logical sense: an act's "intrinsic nature" 'makes' it the "kind" of act it is, in the sense that any act having that "nature" is thereby said to be of that kind. On this understanding of the term, then, the intrinsic nature of act F is whatever about the act follows logically from its being F. In short, its "intrinsic nature" is simply its defining characteristics. On this usage, "formalism" would be the view that the basic criterion of the moral value of an act is something about what follows from its defining characteristics. But it follows from this that every ethical theory is ultimately "formalist." For if a theory could not specify any kinds of acts which are right just because that is the kind of acts they are, obviously it could never specify any act to be right. The reasoning is the same as in the argument for intrinsic values. Thus, for example, suppose that I say that all acts which produce Q are right. Then, consider an act F, where "F" means 'productive of Q.' If Q-producing is what makes an act right, then act F will have to be said to be "intrinsically" right, on pain of nonsense. For act F doesn't, of course, produce Q. In order to say of something that it "produces" something, we must understand the thing produced to be distinct from what is doing the producing. Mills produce flour; but they do not "produce" mills: they are mills. Similarly, acts of Type F do not produce Q—they are Q-producers. Producing Q is their "intrinsic nature" (i.e., their defining property).

It is sometimes supposed that no line can be drawn between acts and consequences. This simply is not true. It is true that the term 'act' can apply to a tremendous range of different things. Sometimes our descriptions of acts, for instance, include a great deal of what would be accounted consequences, relative to another and narrower description of the events in question. For example, Jones starts a revolution. This is an "act"; but perhaps he started the revolution by starting a riot. The riot, perhaps,

\footnote{Cf. Chapter III, Intrinsic Values.}
caused the bringing up of troops, which in turn led to bloodshed, which led to the collapse of the government. Relative to the act of starting the riot, the bringing up of the troops was a "consequence," but relative to the act of starting the revolution, this was part of the act. Given any specific description of an act, what is to count as the range of "consequences" follows: It is merely all of those events, causally connected with the performance of the act, which lie outside the logical boundaries drawn by that description. It is true that no one description is "the" description of "the" act. Did Jones start the riot or the revolution? Well, he did both: the one by doing the other.\(^4\)

As a point of ordinary language, we may observe that we can draw useful distinctions between 'act,' 'circumstance,' and 'consequence.' Roughly speaking, we describe as his 'act' the events which the agent saw himself as producing. Thus, if Jones killed Smith, we call this an act of "murder" if Jones' intention was to kill Smith, and if it was not his intention simply to defend himself from an assault of Smith's, or to carry out a duty as State Executioner, as a soldier, or the like. Let us suppose that Jones hit Smith with a crowbar, in consequence of which Smith died. If we say that Jones murdered Smith, we imply that if, for instance, there had been a wrench instead of a crowbar, he might have used that instead, or might have shot him if he'd had a gun. The specific act he can be said to have "done" is determined by his broader intention. In turn, "consequence" is usually confined to events which are not only caused by the act in question, but also such as could, within reason, have been foreseen by the parties acting, or at least by observers. Circumstances, finally, are part of the "background" of the act, in the sense that they are the conditions which prevailed in the neighborhood of the act. It is not necessary for us to go into these fine distinctions for our purposes. It is enough to point out that some descriptions are ordinarily classified as descriptions of the act, others of the consequences, and others of the circumstances, but that unusual ones are sometimes possible; in any case, that these distinctions are relative, so that to classify a certain description under the "act" category automatically implies that some other

MORALITY AND UTILITY

descriptions will be classified as “consequences,” and perhaps still others will be put into the “circumstance” department.5

Now to apply all this to utilitarianism. Suppose there is a type of activity which, by definition, consists of producing something which is reckoned by utilitarians to be (morally) a good thing to produce. An example is “entertaining,” which by definition produces pleasure or enjoyment (entertainment) in those to whom it is done. According to utilitarianism, enjoyment is “intrinsically” good. It follows (according to utilitarianism) that entertaining is an intrinsically good activity, since its being morally good follows from its intrinsic nature. On the other hand, an “extrinsically” good act is one in which the fact that it is good is not due to its intrinsic nature. Thus, if I happen to bring great joy into somebody’s life by giving him a baseball bat, that is an act whose goodness is not due to its intrinsic nature. I might have given it to him, but he might not have derived any joy from it at all. Some people will dispute whether this was not a “good act” anyway, because of its intention, but this is another issue, quite different from the present one (as we shall see in the following section, Intentions and Acts.) Thus, there may be quite a few types of acts which will turn out to be not only right but intrinsically right, on the utilitarian view. In fact, I shall argue that virtually all of the kinds of acts brought up by formalists as supposed counter examples to the utilitarian theory are of this kind. Examples are promise-keeping, gratitude, and reparation, which turn out to be intrinsically morally good or obligatory; and lying, cheating, stealing, and robbing, which turn out to be intrinsically morally wrong.

Confusion on this question of formalism stems from two sources. The first is the assumption that if an act is intrinsically right, then there cannot be any explanation of why it is right. We have already accepted the argument that if anything is good at all, then something, or some things, must be good in themselves, rather than because of their consequences. This argument seems to me quite undeniable; but it only shows that the kind of explanation we have to give for a thing’s being good in itself must be different from the kind of explanation appropriate to

5 There is a clear and useful discussion of these matters in Eric D’Arcy’s book, Human Acts (New York: Oxford University Press, 1963).
things not good in themselves. It does not show that an explana-
tion of their goodness is impossible. There is no incompatibility
between a teleological view, according to which the ultimate
criterion of a certain kind of goodness lies in certain effects, and
the assertion that some acts are good in themselves. The explana-
tion is simply that some acts logically include the effects in ques-
tion: it is logically impossible to kill someone unless the victim
ends up dead, even though there is only a causal (hence contin-
gent) connection between pulling the trigger and the victim's
death, or between pushing him over the cliff and his dying. It
is a pity that failure to make this elementary logical observation
should cause so much confusion in the ranks of moral philos-
ophers.

I take it as proved then, that utilitarianism (or any ethical
theory whatever), holds that some acts are intrinsically good,
some are intrinsically bad, and some indifferent. So the differ-
ence between formalism and teleology, or deontology and tel-
eology, cannot be properly made out in this way.

The gimmick in this question, as we have seen, is the notion
of "consequence." In effect, this word has been seen to be am-
biguous: on the one hand, it means roughly, "result of something
done by the agent." Elsewhere, I have invented a name for this
distinction. Suppose my act is murdering P, then the conse-
quences are such things as the grief of his widow, setting up of
the hue and cry by the local police, and so forth. But P's dying,
which is logically a part of my having murdered him, is also a
non-logical consequence of something I did—for example, pull-
ing the trigger or shooting him. Relative to the description 'mur-
dered P,' this is part of the act; relative to the description, 'shot
P,' it is a consequence. But shooting P was part of murdering
him. Then, let us call his death an "infra-consequence" of my
murdering P. That is, "infra-consequences" will be those conse-
quences of portions of my activity which are necessary condi-
tions of applying some larger act-description.

If, we were to ask now, in a purely general way, what we
would be left with if we were to try to exclude all consequences
and all "infra-consequences" of acts in a moral theory, we should

---

8 In the article, "Utilitarianism and Formalism" (see n. 1).
find ourselves left with the ghosts, so to speak, of the actions which have always been main concerns of moralists. For there is hardly anything a person does which might not have been "infra-consequential" in the sense in question. Is raising my arm, for example, something I do, or a consequence of something else I do, such as issuing some kind of internal volition? As we know, an answer to this question led Prichard to adopt the view that all we can be required to do is to "set ourselves" to do something, rather than simply to do it. Thus formalism, pursued with fanaticism, leads to "subjectivist" or "intentionalist" views, which we will consider in the next section. It is enough to observe here that the only meaningful controversy there can be between utilitarians and others is whether utilitarians have the right kind of consequences in mind. If we include both "consequences" and "infra-consequences" in the scope of the utilitarian theory—which we obviously must—then the question whether consequences are relevant at all becomes absurd. That is, unless it is metamorphosed into the question of whether, and if so why, intentions are relevant.

Perhaps it is necessary to add a cautionary note at this point, stemming from the liberal employment of such words as ‘definition’ and ‘nature’ in the foregoing account. The language in which ethical rules and principles are formulated is, of course, ordinary language, and all of the appropriate clichés about such language must be duly invoked to avoid an unwarranted assumption of precision. Thus, when I say that I hope to establish, in the cases of nearly all of our "standard" moral principles, that the things said to be right or wrong in those principles are intrinsically right or wrong according to the utilitarian view, we must admit that we aren’t going to be able to say, for instance, that "lying" is, by definition, precisely this or that, and hence . . . . What we do have to say is that the logically normal cases of lying (or whatever) are like this or that, and in those cases, such and such holds, and thus. . . . However, if we take a case

FORMALISM

far enough out on the fringe, perhaps it will be one of those rare cases in which the principle breaks down, but alternatively, perhaps it isn’t really a case of lying. Thus, recall Austin’s remarks about ordering and warning:

He did “order” me to do it, but, having no authority over me, he couldn’t “order” me: he did warn me it was going to charge, but it wasn’t or anyway I knew much more about it than he did, so in a way he couldn’t warn me, didn’t warn me. We hesitate between “He didn’t order me,” “He had no right to order me,” “He oughtn’t to have said he ordered me,” just as we do between “You didn’t know,” “You can’t have known,” “You had no right to say you knew.”

This is, of course, merely by way of general caution. Specific cases will be looked into by turns in due course, and they are what really count.

“Rigorism” and the Status of Particular Moral Principles

It has been said that utilitarianism encourages moral laxity by always making us look for reasons for keeping promises, paying bills, and other obligations. The suggestion is that there are certain rules of morality which will lose their authority if we think of them as subject to exceptions in the light of the overarching importance of the general happiness. A theory like Kant’s, on the other hand, puts its money on real moral rigor—whatever that is supposed to be.

Now, there are some people whose activity is excessively rigid: they are always so anxious to be doing their duty that they walk miles to repay a shortchanged penny, frown whenever anyone tells a risqué story, and always watch suspiciously for signs of moral decline in their neighbors. As Baier says, they “... hunger and thirst after righteousness as others do after food and sex. They are moral automata, in the clutches of their superegos.” Such people raise psychological, as well as moral

---

9 Baier, Moral Point of View, p. 13.
MORALITY AND UTILITY

problems. For our present purposes we have merely to note a by now well-exposed confusion: that is, between the truism that morality comes first, that where duty calls one must follow; and the "falsism," as we might call it, that there are certain types of acts which one must always do, certain specific ethical principles which cannot be overridden by anything. The latter might be called "rigorism." It is necessary to distinguish between this and what I have been calling "formalism" (or "deontologism"), since the one does not imply the other. A formalist holds that we ought to do something because of its "intrinsic nature," and not because of its "consequences"; it does not necessarily follow that if an act has a nature which is such as to render it a duty, we must always perform it when the circumstances make it possible to do so. In fact, as the previous analysis showed, to say that something is "intrinsically right" is equivalent to saying, in effect, that it is "prima facie" right. By now, I take it to be commonplace that it is one thing to say that a certain kind of act is one's duty, and quite another to say on a particular occasion that this act is your duty. Baier has put this very well in distinguishing what he calls "reasons prima facie" from "reasons on balance." In any particular situation, we are faced with choices of action. We consider various acts, in each case noting that it is of various kinds, i.e., that each act has various characteristics. Some of these characteristics will be such as to count in favor of doing it, and some against. Then we have to decide whether the reasons for doing it outweigh the reasons against it. If the characteristics in question are such as to make for reasons of a moral kind, and in particular of a kind having to do with duties, then it may be that, taking all things into account, the act is a duty. Some of the characteristics will be "duty-making" (i.e., their presence is such as to suggest that there is a duty to do the act), but others may be such as to suggest there is a duty not to do it. In that case, these characteristics have to be weighed, and the result will be one's duty in the given situation. But there is no specific kind of act which it is always right to do, regardless, or always wrong to do, regardless. It is conceivable that

10 The importance of such a distinction was emphasized also by Ross in The Right and the Good, but his account is not as clear as Baier's, and is linked too closely with Ross' brand of "intuitionism." Baier's account is found in Moral Point of View, pp. 102, 103.
FORMALISM

any two particular moral principles might conflict. For example, we may find that in order to avoid stealing, we must kill, or that in order to help a friend, we must harm an employer. "Rigorism" of Kant's kind is impossible in these cases. Kant, as is well known, felt that if a particular kind of act is a duty, then it will always be a duty, come what may. He failed to see the distinction between "prima facie" duties, and duties "on balance."[11]

However, if one holds that there is only one ultimate moral principle, then it cannot be merely prima facie. In view of the suggestions in Chapter II, if an act is a moral duty on the whole, then it cannot "have exceptions"; and if there is but one moral principle, and it makes a certain act a duty, then there is no further recourse. It follows, then, that the Utilitarian, since he is a "monist" in this sense of claiming that there is but one ultimate moral principle, is necessarily a "rigorist" with respect to that principle. The typical "formalist," on the other hand, claims that there are a number of ultimate moral principles, so he cannot be a rigorist with respect to any particular one of them.

We can produce a general proof of this last point. To say that there are several ultimate moral principles is to say that there are several logically independent predicates, call them \( F_1, F_2, \ldots, F_n \), each of which figures in a statement to the effect that for any act, if the predicate in question is true of it, then it has some moral value, \( V \). Presumably, a moral system, to be nonvacuous, must have some principles specifying what is wrong as well as others specifying what is right; i.e., it must specify some things which ought to be done, and others which ought not to be done. Then, let us take any two of its principles, say those involving \( F_1 \) and \( F_2 \), for convenience, such that they specify opposite values for the acts respectively instantiating

[11] Baier, Moral Point of View, 191–95. Baier makes another distinction between "presumptive" and "prima facie" reasons, which, he says, Kant overlooks. I shall consider a similar distinction in the chapter on Justice, below. According to Herbert J. Paton in Kant-Studien (Cologne: 1953–54), Vol. 45, Kant wrote the famous piece in which this doctrine is most explicitly defended in a fit of pique. Nevertheless, he did publish it, and though it is inconsistent with some of his practice, it is quite consistent with many of his statements in the well-known Kantian works on ethics. The small essay in question, "On a Supposed Right to Lie from Benevolent Motives," is found in an Appendix to Thomas K. Abbott (tr.), Critique of Practical Reason, and other works on the Theory of Ethics (London: Longmans, 1909).
the predicates in question. The formulas of these principles will be \( (x) \ (F_1 x \supset V x) \) and \( (x) \ (F_2 x \supset V_1 x) \) (where \( V_1 \) is some value incompatible with \( V \)). It follows that any act which is both \( F_1 \) and \( F_2 \) will be given incompatible values. Therefore, the system will lead to contradictions in some cases. Thus, at least most of the principles in the system cannot be held "rigorously" in the above sense, and under ordinary circumstances, none of them will. Most of them will have to be, as Ross saw, "prima facie" principles. It might be held that no act can ever have both of any two predicates whose correlated values are inconsistent, but this cannot be held on logical grounds, for to say that \( F_1 \) and \( F_2 \) are logically independent is to say that \( F_1 x \cdot F_2 x \) is consistent. The compatibility of principles in application can never be assured, except by reducing the number of principles to one. (Or by cheating, e.g., by saying that everything is good.)

Given a particular moral code, of course, it will be possible to live up to it well or badly. A person who lives up to it perfectly might be said to be a "rigorist" with respect to that code. But it is plain that "rigorist" is not a theoretical term at all as used in this way. Obviously, we ought to do whatever is right, and if a rigorist is simply someone who always does what is right, then we ought all to be rigorists. But this has nothing to do with moral theory. One can be a rigorous utilitarian or a lax Kantian, or vice versa, in this sense of the word. Objections to utilitarianism on this score are merely ad hominem.

It may be felt that even though we are not obliged always to keep promises, utilitarianism provides for more exceptions to them than it ought. This may be one source of the allegation that utilitarianism "cannot account for justice." "Rule-utilitarianism" is in part an effort to shore up the status of particular moral rules without altogether abandoning utilitarianism. I will deal with this in Chapter V.

**Intentions and Acts**

We commonly apply moral terms to intentions and motives, as well as to intentions with which they were performed. (This
is not surprising, since ‘act’ is basically an intentional term in
the first place.)

Some have gone so far as to hold that only motives or inten-
tions have genuine moral worth, with Kant, of course, as the
principal case in point. This is understandably thought to be
a great difficulty for the utilitarian, and it is sometimes thought
to be the single blow that fells the whole theory. If the moral
value of an act is proportional to its utility, then how can the
intentions or motives behind it have any bearing on the matter?
In the present section I shall argue, as usual, that there is a good
deal of confusion surrounding this question. I shall further argue
that the view that intentions are the basic bearers of moral pred-
icates is incoherent. In the following section, I shall go on to
argue that only a utilitarian point of view makes good sense
of all the facts involved here.

The problem then, is as follows: If we are to take account
only of the consequences of acts and not their intentions, then
how can we account for the fact (as we may call it) that an
act which turns out unfortunately but was well intended is more
excusable than one in which bad effects were intended? Notice
that the objection presupposes that the intentions of an agent
could be appraised quite independently from the acts to which
they are “directed.” If it could be shown that this is not so,
then we shall be well on the way to showing that the objection
is incoherent. That is precisely what I propose to do.

Let us begin by recalling the primary functions of ethical
principles: to tell us what to do, i.e., to guide action. Whatever
else an ethical principle is supposed to do, it must do that, other-
wise it could not (logically) be an ethical principle at all. Now
let us consider the way in which we can appraise a person’s in-
tentions. In order to do this, clearly it has to be possible for
some intentions to be better or worse than others: our criterion
for appraising them, that is, must be such as to distinguish among
them. In order to do this, it must make use of some of the aspects

Kant usually says that only the “will” can be called good “without
qualification”; an action which is done “from duty” derives its moral worth
from the maxim of the act, and “therefore” from the “principle of voli-
tion.” This, as well as other passages, strongly suggests that the moral value
of acts derives from the moral value of the will behind them, and not vice
versa. This is the issue which is discussed in this section. Although most
interpreters would agree with Kant’s view, I claim that it is upside down.
in which they can differ from each other. And if one reflects on it, it will shortly become evident that there are, in fact, only three ways in which intentions can differ. They can differ, namely, in their strengths, in their "sincerity" (if this is different from "strength") and in their "objects," that is, in what they are intentions to do: i.e., in the kind of act which will occur if the intention is carried out.

Now neither of the two first are differences which can enable us to rate intentions morally. For plainly, a person can have a very strong evil intention, and also a very sincere one. He may, for instance, be beating a heretic at a stake, and he may really and sincerely intend to do this. It follows that neither of these will do. Since there is only one left, it must be the one we need. We must, then, appraise intentions by reference to their "objects," i.e., by what they are intentions to do. But it is logically impossible to do this without having criteria for rating the acts which intentions are directed at. Thus, criteria for rating acts independently of their intentions are presupposed in any effort to appraise intentions for their moral worth. The worth of an intention is not, and cannot be, due to anything else but the value of the action intended (i.e., which the action would have, if it occurred as intended.) It is this which the Principle of Utility provides: a criterion for evaluating acts quite apart from the intentions of the people who perform them.

Someone will say, but you have not yet shown us how it follows from the utilitarian position that we should rate intended acts by the utilitarian criterion, and not just acts apart from intentions. How can we do this? However, the foregoing argument also shows that the Principle of Utility does provide criteria for rating intentions. For, if we can rate acts apart from their intentions, then we can also rate intentions by appraising not the act which actually was carried out by the agent, but rather the act which he intended to perform, whether or not

13 Oddly enough, some of the "existentialists" suggest that sincerity alone is sufficient. Sartre seems to be an example. "Authentic" living, with which the existentialists seem much concerned, appears to be the same as sincere living, although I suppose that this will not be thought profound enough by most of Sartre's adherents. Cf. Jean-Paul Sartre, Existentialism and Humanism (London: Methuen Publishing Co., Ltd., 1948), especially pp. 50-51.
he succeeded. Now I must show not only that we can do this, but also that the principle of utility implies that we must.

At this point, it is again necessary to remind the critic of something which he often seems to forget when he discusses the utilitarian theory. That is, the purpose of any moral principle. Utilitarianism, as I have insisted, is a fundamental moral principle. Moral principles, it is agreed, are for the purpose of guiding action. The reason why utilitarianism implies that we should appraise intentions as well as acts follows logically from this. It is simply nonsensical to pose the question, "ought I to intend to do this, or ought I merely to succeed at it?" Obviously, I can have no such choice. All I can do is to seek guidance, and then proceed to follow it to the best of my ability. Conversely, when I seek to influence someone else's behavior by offering him advice, I am, ipso facto, attempting to alter his intentions. That is what it is to offer advice. If I thought that advice would not do him any good, and it appeared that he intended to do evil anyhow, I should have to resort to force or to psychological trickery to bring him around. Obviously, it is sometimes necessary to do this. But that is to depart from the making of ethical statements, which is our primary concern here.

In short, this famous and ancient objection to utilitarianism is simply misguided. The answer to it is parallel to the answer provided by Baier\footnote{\textit{Moral Point of View}, pp. 143-47. Cf., also the following section of this chapter, entitled "Objective" and "Subjective."} to the ancient problem about the "paradox of subjectivity": ought a man to do what he thinks he ought to do, or what he really ought to do? The answer is that he has no option but to try to figure out what he ought to do, and then to do it. If his deliberation is faulty, we can blame him for that; if he fails to do what he agreed he ought to do, we can blame him for that. But there is nothing else left to blame him for, and these are different shortcomings.

Actually, utilitarianism does better on this subject than merely to escape this ancient objection. It is, in fact, the only theory that makes sense of all the phenomena involved. For example, consider a theory such as Kant's, according to which the intention is the only thing that counts. We have already seen that this is logically incoherent, if interpreted so as to imply that
MORALITY AND UTILITY

criteria for appraising acts are unnecessary. Plainly Kant could not have meant that, since his Categorical Imperative enables us to appraise acts apart from their intentions anyhow. The Categorical Imperative enjoins us to act only on maxims which could be universalized; but people can act as if they had a certain maxim even if they don't have it, and such actions can be appraised by reference to the principle that if they were performed with that maxim, then they would be wrong. It would be nonsense to interpret this merely as enjoining us not to act with that maxim (though acts of the type which the maxim suggests are all right, so long as they are not done with the maxim in mind). On the contrary, the only point there could be in enjoining us not to have that maxim is not to act on it. This implies that it is really the act which, in the logically prior sense, is bad. We can admit, of course, that we might want to say that unintended acts of the proscribed type should be called "unfortunate" rather than "wrong," but the point is that we can use the Categorical Imperative's test to determine the class of unfortunate acts as well as that of wrong ones, and it would be quite incoherent to maintain that it can be used only to determine that of wrong ones, but not unfortunate ones.

What Kant may have meant is that regardless of how frightful the results of what a man actually does, he cannot be blamed so long as he meant well. If this is what he meant, then clearly his doctrine is false. A man who intends to drop an H-Bomb on New York City but who "means well" by it is not to be praised (supposing that he is, in fact, unjustified in this, which is likely supposition). He should be locked up, or at very least, put under observation. What is true is that the kind of treatment which should be given to a man who knew that blowing up New York City is wrong, but who did so anyhow just for fun, and

15 This is strongly suggested by the famous passage at the beginning of the *Fundamental Principles of the Metaphysics of Morals* (tr.) T. K. Abbott (New York: Liberal Arts Press, Bobbs-Merrill Co., Inc., 1949), p. 12: "Even if . . . this will should wholly lack power to accomplish its purpose, if with its greatest efforts it should achieve nothing, and there should remain only the good will . . . then, like a jewel, it would still shine by its own light, as a thing which has its whole value in itself." Unfortunately, Kant does not really face the problem of what to say when it not only accomplished nothing good, but through bungling, accomplished a lot of harm. Cf. Nowell-Smith's fine discussion of this in *Ethics*, Ch. 17.
the kind of treatment to be given to a man who thinks that New York City is Sodom and Gomorrah incarnate and that the world would be morally cleansed by its disappearance, are quite different. But why should they be treated differently? I do not see how this can be answered intelligently, except by considering the probable outcomes of the different kinds of treatment.\textsuperscript{16}

The fact, of course, is that all of our acts can be appraised by the principle of utility, including the acts of praising and blaming, of punishing, and of deliberating themselves. It would make no sense to blame people for what they did not intend to do, because "blaming" them is not going to enable them to avoid mistakes in the future. Only if it is going to do this, can we properly blame them.\textsuperscript{17} Consider, for instance, a three-year-old child who does something he does not know is bad. If one blames him for doing so, it may well cause him to reflect before doing so in future; and if so, then the blame is to that degree justified. The same is true of people who are always going about doing evil with good intentions. Blame may well make them realize that something is wrong, and in the extreme case, it could cause them to adopt the principle, "Whenever I have good intentions, I should avoid acting on them." This policy might save them and everyone else a lot of trouble, in some cases. I do not see how any other view than a utilitarian one can account for the fact that we realize that such things do happen.

\textit{"Objective" and "Subjective"}

It is plausible to say that 'wrong' is sometimes used in a sense equivalent to 'morally blameworthy' or 'reprehensible' in a sense which implies the propriety of disapproval of the agent for his deed.

\textsuperscript{16} Further discussion of this is found in the following section; also Ch. VI ff.

\textsuperscript{17} This has been noted by Bentham, Mill, Sidgwick, and Smart, among others. Brandt's recent article, "Toward a Credible Form of Utilitarianism, in Castañeda-Nakhnikian (ed.), Morality and the Language of Conduct, will be discussed further below. It is surprising how little they make of it; and it is unfortunate that in the case of Smart, for example, the use made of it is sometimes wrong-headed. Smart holds that a utilitarian might be justified in advocating some other moral theory on the ground that the time isn't ripe for its reception. (Cf., Outline, p. 36, for example.) Perhaps this is not wrong-headed, but it is certainly misleading.
MORALITY AND UTILITY

Now, if utilitarianism is understood as a theory about right and wrong actions in this sense, I believe it is indefensible in all its forms. For we have good reason to think that whether an act is wrong in this sense depends in part on such things as whether the agent sincerely believed he was doing his duty, whether the temptation to do what he did was so strong that only a person of very unusual firmness of will would have succeeded in withstanding it . . . If whether an act is wrong depends in part on any one of these factors, then it is hard to see how the utilitarian thesis that rightness or wrongness is in some sense a function of utility can be correct.\(^\text{18}\)

This passage by R. B. Brandt may be taken as typical of this type of criticism of the utilitarian view. Brandt goes on to suggest that a utilitarian-like view, which he elaborates in the essay from which the above is taken, may nevertheless be acceptable as a theory of right and wrong in what he calls the “objective” sense. It is this sense, I have argued, that one must employ when considering what to do (and which is the basic sense for ethical predicates). When one is trying to make a decision, one is not in a position to “blame” or “praise” himself for what he is about to do, because that is what he is trying to decide. Neither can he excuse himself for having an inadequate sense of right and wrong, because he is supposed to be exercising it at that time, and then, if he suspects himself of an inadequacy, to be trying to rectify it; he cannot excuse himself because he was provoked, since his question is whether or not to respond, and if so, in what way, to whatever provocations there are.\(^\text{19}\) According to the utilitarian view, he ought to determine which of the actions open to him would produce the most value, as measured by utility in the sense we developed in Chapter III, and then do that (subject to the qualifications elaborated below). One could appraise the value of an earthquake, or an accident, in the same way, of course. One earthquake is worse than another if it kills more people than the other, causes more suffering and property damage, and so on, just as the Nazi regime was worse than the Mussolini regime because it caused the deaths of more people, more suffering, more property damage, and so on.

Now, whether there are two senses of ‘wrong,’ as Brandt's


\(^{19}\) Sartre’s “Existentialism” seems largely an inflation of this point.
passage implies, is a difficult and disputed question. It is certainly true that we make appraisals of right and wrong for different purposes, and that such appraisals seem inappropriate in some contexts, but not in others. For example, it would be odd to say that an earthquake did something "wrong," even if its effects were the same as those of a crime wave, where we would say that the criminals had indeed done wrong. This certainly suggests that the sense of blaming is intimately connected with the use of 'wrong'; but it also suggests that there isn't any "objective" sense of 'wrong' at all. On the other hand, 'good' and 'bad,' 'worse' and 'better' apply equally to earthquakes and human acts. This suggests that the vocabulary of evaluation (as Sesonske calls it), or the "teleological" vocabulary (as Nowell-Smith calls it), could be reserved for the "objective" sense and the vocabulary of "obligation," or the "deontological" vocabulary (as they are respectively called by the same authors), could be reserved for the "subjective" sense.

But this will not quite do, either. When we appraise characters, as opposed to acts, plainly the "subjective" considerations about intentions and so forth are relevant, and yet 'right' and 'wrong' don't apply while 'good' and 'bad' do. There is nothing for it but to proceed slowly, and to explain each set of cases in turn, without hoping too much for a systematic vocabulary.

To begin with, there is obviously a connection between "subjective" and "objective" uses. In the ordinary course, if I say to Jones, "That was very wrong," in reference to his action, I am indeed blaming him; but, I am blaming him because of what he did. Obviously, my sentence on such occasions normally has both the "objective" and the "subjective" senses. Saying to Jones that what he did was wrong blames him, and indicates that there was a reason for blaming him, viz., a reason of a moral kind. Moreover, the fact that what he did was wrong "objectively," if it occurs to Jones retrospectively, will very often of itself constitute a retrospective blaming. It does so when the agent realizes that he could have avoided the act if, for example, he had been thinking what he was doing at the time.

Whether or not we should call these two functions of ethical

---

21 Nowell-Smith, Ethics, p. 223 and elsewhere.
terms distinct "senses" of those terms, we can at least give a utilitarian account of why the difference is called for. The reason is not far to seek. The clue to the matter is that the principle of utility may be used to appraise any act whatever, including acts of praising and blaming, criticizing and censuring. Now, the "subjective" senses or functions of ethical terms in which, as Brandt says, they "imply blame," are clearly performative. If I tell somebody, "What you did was wrong," I am (normally) blaming him, and this is an act of mine. What considerations, now, are relevant in trying to decide whether to blame him? In particular, whether to blame him, as opposed to non-emotively informing him that he could have done better (for instance), or keeping quiet and forgetting it? I propose that we can use the principle of utility to decide this, and that doing so will show why considerations about the person's intentions, the state of his knowledge, the degree of temptation to which he was subjected, the degree of psychological pressure under which he was operating, and so forth, are relevant.

The basic consideration in all this, as urged above, is that the purpose of ethics is to guide action. What does this mean? Consider the difference between getting someone to do something by giving him a drug, by hypnotizing him, by threatening to put him in jail if he doesn't, by offering him a reward if he does, by commanding him to, by telling him to, by requesting him to, by suggesting that he do it, or by reasoning him into doing it. These are all different ways of controlling or influencing behavior. Corresponding to these differences, there are differences in the subjects whose behavior we are trying to control. (Notice that it might be our own; I am speaking generally here.) If a person is in a foaming rage, the last several methods are unlikely to be available. On the other hand, if he is perfectly calm, knows just what he's doing, and speaks your language, then several of the first methods listed will be inappropriate. Why? Because they cause the subject much more trouble than the later ones, in general, and also because they are, in a sense, not likely to work. Obviously, the Principle of Utility calls for the use of that method which will cause the subject the least amount of harm, discomfort, pain, damage, loss, etc. If he is a reasonable man, and I request him to do something, then he will do it if it's a reasonable request. Suppose that it's a request which I believe
to be sanctioned by the principle of utility. If he regards the principle of utility as a reasonable principle, then if he agrees that it is sanctioned by that principle, he will want to perform the act in question. In that case, resorting to force, blame, or other means are unnecessary; therefore, their use would be wrong.

Now, consider a person who has just done something wrong, and suppose that we have reason to suspect that he didn’t know it was wrong. We should point out to him that it was wrong, so as to put him on guard against doing that kind of thing in future. Blame is out of place, because it’s more than is needed (so far as we can see) to correct his behavior. Suppose, however, that though he knew it was wrong, he couldn’t help it. In this case, blame is out of place, but not because it’s more than is needed: it’s because it wouldn’t do any good. If he couldn’t help it, then in future cases of the same sort he won’t be able to help it either, and the fact that his ears are ringing with the scolding we gave him last time will not succeed in preventing him from doing it again. This mental pain will be quite useless. What we need to do is to go into the circumstances and try to remove the compelling factors. Thus, if he did it because Al Capone was threatening him with death, we should go out and round up Al Capone. If he did it because there was a landslide and he was making a frantic effort to escape, we shall advise him to move to a safer area or to put up a retaining wall, or we shall conclude that it was just an unfortunate accident.

Suppose he did it because he thought he was perfectly justified, though in fact he wasn’t. Now, a utilitarian is someone who believes that actions are justified by their utility; obviously, he is not someone who believes that we ought not to perform actions that are justified. We are not saying, “Never mind whether the act is justified: look only to its utility.” This would be perfectly perverse. Thus, if the individual thinks that it is justified, and that is why he did it, then there is no sense in blaming him,

---

22 Cf., for instance, H. D. Aiken, “Moral Reasoning,” in *Reason and Conduct*, p. 100. He remarks that the “ordinary utilitarian would have to disallow the claims of justice except insofar as the obligation not to act unjustly is derivable from that to maximize happiness or reduce misery.” This is highly tendentious, since it carries weight only if the utilitarian can’t derive the whole of the obligations of justice from his principle. Aiken, like so many others, treats it as self-evident that the utilitarian cannot do this. But I propose to do just this, in Ch. VI below.
just like that. For we want people to do what is justified, and he thought he was justified. What we have to do is to consider whether to blame him for thinking that what he was doing was justified. And this is quite a different matter from blaming him for what he did. It might be a great deal less, or it might be a great deal more than that. If the act was wrong, but not seriously wrong, whereas his justification was seriously, horribly wrong, then blaming him for thinking badly in a fairly strong tone of voice might be called for. It is called for because that type of thinking could lead to much worse actions in future.

However, suppose that although his ethical thinking is shocking, nevertheless there has been a lot of it. He is, let us imagine, a thoughtful but misguided Existentialist, or even a theorist for some fantastic fascist “ethics.” This is an extreme case, and there might be no hope; but if there is hope, then certainly the proper thing to do, if we have the time and means, is to sit down and reason with him. For if utilitarianism is a truly reasonable philosophy, and his is really unreasonable, then we ought to be able to show him this, if his mind is in working order. We shall have to go into the material of Chapter IX with him, perhaps.

In brief: blaming people, criticizing them, ordering them, and the like, are all ways of influencing people’s courses of action. They are all, therefore, means which should be avoided unless justified by overriding utilities. On the Principle of Utility, it follows that we should use the method which causes the least harm in the form of mental or physical distress. Therefore, we should inflict just that degree of blame which is necessary to get the individual’s behavior in line, and the kind of blame which will do the job. Moral praise and blame occupy a limited, but extremely important, spectrum of the methods of influencing people’s behavior. What is limited about them and what is important about them are due to the same thing, namely, that they are verbal, “rational” methods. If I accost someone with words in order to change his behavior, I am using a method which won’t work unless he is rational to some degree. There is no use blaming mountains, because they don’t understand the concept of blame (nor anything else). There is a kind of “blame” which

FORMALISM

works for horses and dogs, but not for mosquitos; and there is a kind of blame which works for people, because they are more intelligent than horses and dogs. Finally, we get to the point where a mere remark or proposition uttered without heat or fury, and indeed, one that gives something close to pleasure in a rational agent, is quite sufficient. Being able to sit down and coolly reflect on one's own behavior, without great agonies of conscience, indulgence in neurotic self-reproach or "existential anxiety," and still to effect important changes in one's future behavior by that means, is the mark of the rational being from the point of view of practice. The language of ethics is the language which such a person employs at its best. The principle of utility explains why we should employ that language in appraising the conduct of fully rational people who know what they are doing, whereas we must resort to successively harsher or cruder or more psychologically dynamic methods with other people, or, in the extreme case, with animals or even inanimate objects. The latter we just push around without a care for what they think, because they don't. The former are beings we do care for, morally, and therefore we want to use the least unpleasant methods with them.

This account also shows how absurd certain of the controversies about "free will" become in ethical contexts. The reason for using one or another kind of blame, in reference to this or that aspect of a person's activity, has to do with the likelihood that it will, in fact, bring his behavior into line with the requirements of morality, as opposed to the amount of disutility their use will cause. This involves consideration of "responsibility." At no point is there any sense in trying to determine whether the subject's behavior—and remember, this includes one's own behavior when one is appraising it—was or was not "caused" at all.  

Indeed, no decision on the subject of whether human behavior is predictable in principle or not, or whether some of it is gov-

24 See Smart, *Outline*, p. 39. "The 'Act' utilitarian will quite consistently reply that the notion of the responsibility is a piece of metaphysical nonsense and should be replaced by 'Whom would it be useful to blame?'" It is better to say: The notion of who is responsible is the notion of whom it would be useful to blame. This view is well argued by Nowell-Smith, *Ethics*, Chs. 19–20.
MORALITY AND UTILITY

erned by chance, is required for ethical purposes. In every case, we have only to look at the facts. Some behavior is apparently random and unpredictable (so far as we can see, at present), and one has to allow for this. We can, by the way, "predict the unpredictable," to put it paradoxically. Certainly many or most of us have had the experience of setting out to do one thing and finding ourselves doing quite another, and learning to avoid situations where we might find ourselves behaving quite randomly or irrationally, if the random or irrational behavior in question is likely to lead to harm. It is equally obvious that what a person thinks about his behavior often makes a difference to it. This has nothing to do with "metaphysics," but is a matter of plain observation, as Locke and Hume correctly pointed out.

This last point, incidentally, is the reason why fantasies about universes of people who are always trying to do the right thing but always fail are indeed "fantastic." Beings who are unable to control their actions in the light of their beliefs cannot develop a language of ethics, for then it would be completely unteachable. If people did not frequently succeed in altering their behavior in the light of their beliefs, then the point of having a language of ethics would be lost, and they would be just as well off not having one as having one. In that circumstance, they not only would not try to develop one (if they were capable of "trying," which they obviously would not be), but they would be perfectly justified in not developing one.

Implications for the theory of punishment are readily available from the above account, and fortunately, are to some extent being put into practice today as well. In particular, this account emphasizes the continuity between procedures of criminal justice, such as incarceration, fining, probation, and capital punishment, and those of the psychiatric profession, right on down to abstract moral philosophy. A further advantage is the naturalness with which an account of the role of rules follows. The precise way in which rules are justified, the degree and limits of the justification, and the type of rules which can be justified, will be discussed below. But the general consideration is clear: a rule is promulgated, taught, or otherwise put into circulation

25 Nowell-Smith, Ethics, p. 250.
in order to enable people to anticipate (and keep track) of occasions when blame would be in order and when it would not. If, as I have urged, the essential feature of the language of morals is that it is designed to influence the behavior of rational beings by giving them reasons for various courses of conduct, then the purpose of such language is defeated if a rational agent cannot predict when predicates of praise and blame will be applicable and when they will not. If we were all calculating machines and could appraise our actions in the light of the Principle of Utility with no reference to rules, then moral rules would indeed be useless. Later, I shall show that this does not apply to rules of games, contracts, implementational rules, and many others.

To some theorists, talk of ethical predicates as essentially stimulative or performative is repugnant, or even appears absurd. Of course, it is silly to suppose that in holding some general moral belief (e.g., that murder is wrong), I am here and now addressing commands, injunctions, and criticisms, to all murderers and would-be murderers past, present, and future, and in all portions of the universe.26 This does not mean that an essentially different account of the uses of these expressions is called for. Surely what we mean when we use these expressions in general formulae is simply that on all occasions when anyone is confronted with an opportunity to issue such injunctions, criticisms, blame, and so forth, he would be justified in doing so. In short, that we have reason to do these things. But reasons for doing some of them are different from those for doing others, as I have taken pains to point out in the foregoing account. In some circumstances, the appropriate thing to say is “That would be very wrong; don’t do it!”; in others, “That’s O.K., but wouldn’t this be even better?” The meaning of a general moral principle, therefore, is logically connected with the meaning of particular second-person uses on specific occasions. And since all evaluative language, as theorists have recently argued, is reason-backing language27 (i.e., carries with it the implication that there reasons

27 See Baier, Moral Point of View. Also, A. E. Murphy, The Theory of Practical Reason (La Salle, Ill.: Open Court Publishing Co., 1965).
licensing the use of the particular performative being employed), it follows that this relation is symbiotic. Use on particular occasions of moral predicates in a performative manner depends for its logical legitimacy on the supposition by the user of the existence of general reasons; but the general reasons, in turn, are reasons for the corresponding stimulative use of language.

I would argue for the further link between the stimulative use of language and stimulation of a non-verbal kind, by actual use of force. The strongest ethical predicates, I shall be suggesting below, may well have their force just because they suggest the appropriateness of the use of punishment or forcible restraint or stimulation. A rational agent can react to the verbal indication that punishment would be in order just as much as to an actual punishment; and indeed, the measure of his rationality as a practical agent is the degree to which (given that he allows the reasonableness of the particular moral judgment being addressed to him, whether by himself or others) he is able to react to these intellectual proddings without having to wait for physical proddings to set him in motion (or restrain him from moving).

Once we understand the essentially performative nature of ethical language, on the one hand, and its connection with reasons on the other, we are in a position to understand how the utilitarian theory is not confined to giving merely an account of the “objective” sense of ethical terms. Indeed, we are able to see why it is misleading to think of the “objective” and the “subjective” as essentially different meanings or senses of ethical words. Unless we make the mistake of literally identifying the “objective sense” of ethical expressions with their so-called “cognitive meanings” (and thus of lapsing into just that kind of naturalism which it was a chief accomplishment of ethics in the first half of this century to refute), we shall have to admit that the differences of dramatic situation due to our remoteness or propinquity to the occasions when direct stimulatory usage is possible do not alter the basic meanings of these expressions. The relevance of “subjective” factors of the kind we have been

---

discussing becomes clear when one reflects on these same stimulative or “performative” functions of ethical language.

In sum, it seems to me that the scruples of Brandt and others on the supposed inability of utilitarianism to account for the “subjective” senses of ethical terms are unjustified. They are due to failure to distinguish between appraisals of the acts to which moral predicates are in the first instance applied, and appraisals of acts of criticizing, blaming, praising, and so forth. Once we make this important distinction, it rapidly becomes evident why the agent’s intentions, motives, state of information, psychological pressures, and so on, become relevant on the utilitarian view.