MORAL RELATIVISM DEFENDED

M Y thesis is that morality arises when a group of people
reach an implicit agreement or come to a tacit under-
standing about their relations with one another. Part of what I
mean by this is that moral judgments—or, rather, an important
class of them—make sense only in relation to and with reference
to one or another such agreement or understanding. This is
vague, and I shall try to make it more precise in what follows.
But it should be clear that I intend to argue for a version of what
has been called moral relativism.

In doing so, I am taking sides in an ancient controversy. Many
people have supposed that the sort of view which I am going to
defend is obviously correct—indeed, that it is the only sort of
account that could make sense of the phenomenon of morality.
At the same time there have also been many who have supposed
that moral relativism is confused, incoherent, and even immoral,
at the very least obviously wrong.

Most arguments against relativism make use of a strategy of
dissuasive definition; they define moral relativism as an inconsis-
tent thesis. For example, they define it as the assertion that (a)
there are no universal moral principles and (b) one ought to act
in accordance with the principles of one's own group, where this
latter principle, (b), is supposed to be a universal moral prin-
ciple.[1] It is easy enough to show that this version of moral rela-
tivism will not do, but that is no reason to think that a defender
of moral relativism cannot find a better definition.

My moral relativism is a soberly logical thesis—a thesis about
logical form, if you like. Just as the judgment that something is
large makes sense only in relation to one or another comparison
class, so too, I will argue, the judgment that it is wrong of some-
one to do something makes sense only in relation to an agreement
or understanding. A dog may be large in relation to chihuahuas

---

[1] Bernard Williams, *Morality: An Introduction to Ethics* (New York, 1972),
pp. 20-21; Marcus Singer, *Generalization in Ethics* (New York, 1961), p. 332.

but not large in relation to dogs in general. Similarly, I will argue, an action may be wrong in relation to one agreement but not in relation to another. Just as it makes no sense to ask whether a dog is large, period, apart from any relation to a comparison class, so too, I will argue, it makes no sense to ask whether an action is wrong, period, apart from any relation to an agreement.

There is an agreement, in the relevant sense, if each of a number of people intends to adhere to some schedule, plan, or set of principles, intending to do this on the understanding that the others similarly intend. The agreement or understanding need not be conscious or explicit; and I will not here try to say what distinguishes moral agreements from, for example, conventions of the road or conventions of etiquette, since these distinctions will not be important as regards the purely logical thesis that I will be defending.

Although I want to say that certain moral judgments are made in relation to an agreement, I do not want to say this about all moral judgments. Perhaps it is true that all moral judgments are made in relation to an agreement; nevertheless, that is not what I will be arguing. For I want to say that there is a way in which certain moral judgments are relative to an agreement but other moral judgments are not. My relativism is a thesis only about what I will call "inner judgments," such as the judgment that someone ought or ought not to have acted in a certain way or the judgment that it was right or wrong of him to have done so. My relativism is not meant to apply, for example, to the judgment that someone is evil or the judgment that a given institution is unjust.

In particular, I am not denying (nor am I asserting) that some moralities are "objectively" better than others or that there are objective standards for assessing moralities. My thesis is a soberly logical thesis about logical form.

I. Inner Judgments

We make inner judgments about a person only if we suppose that he is capable of being motivated by the relevant moral considerations. We make other sorts of judgment about those

who we suppose are not susceptible of such motivation. Inner judgments include judgments in which we say that someone should or ought to have done something or that someone was right or wrong to have done something. Inner judgments do not include judgments in which we call someone (literally) a savage or say that someone is (literally) inhuman, evil, a betrayer, a traitor, or an enemy.

Consider this example. Intelligent beings from outer space land on Earth, beings without the slightest concern for human life and happiness. That a certain course of action on their part might injure one of us means nothing to them; that fact by itself gives them no reason to avoid the action. In such a case it would be odd to say that nevertheless the beings ought to avoid injuring us or that it would be wrong for them to attack us. Of course we will want to resist them if they do such things and we will make negative judgments about them; but we will judge that they are dreadful enemies to be repelled and even destroyed, not that they should not act as they do.

Similarly, if we learn that a band of cannibals has captured and eaten the sole survivor of a shipwreck, we will speak of the primitive morality of the cannibals and may call them savages, but we will not say that they ought not to have eaten their captive.

Again, suppose that a contented employee of Murder, Incorporated was raised as a child to honor and respect members of the "family" but to have nothing but contempt for the rest of society. His current assignment, let us suppose, is to kill a certain bank manager, Bernard J. Ortcutt. Since Ortcutt is not a member of the "family," the employee in question has no compunction about carrying out his assignment. In particular, if we were to try to convince him that he should not kill Ortcutt, our argument would merely amuse him. We would not provide him with the slightest reason to desist unless we were to point to practical difficulties, such as the likelihood of his getting caught. Now, in this case it would be a misuse of language to say of him that he ought not to kill Ortcutt or that it would be wrong of him to do so, since that would imply that our own moral considerations carry some weight with him, which they do not. Instead we can

5

only judge that he is a criminal, someone to be hunted down by the police, an enemy of peace-loving citizens, and so forth.

It is true that we can make certain judgments about him using the word "ought." For example, investigators who have been tipped off by an informer and who are waiting for the assassin to appear at the bank can use the "ought" of expectation to say, "He ought to arrive soon," meaning that on the basis of their information one would expect him to arrive soon. And, in thinking over how the assassin might carry out his assignment, we can use the "ought" of rationality to say that he ought to go in by the rear door, meaning that it would be more rational for him to do that than to go in by the front door. In neither of these cases is the moral "ought" in question.

There is another use of "ought" which is normative and in a sense moral but which is distinct from what I am calling the moral "ought." This is the use which occurs when we say that something ought or ought not to be the case. It ought not to be the case that members of Murder, Incorporated go around killing people; in other words, it is a terrible thing that they do so.[2] The same thought can perhaps be expressed as "They ought not to go around killing people," meaning that it ought not to be the case that they do, not that they are wrong to do what they do. The normative "ought to be" is used to assess a situation; the moral "ought to do" is used to describe a relation between an agent and a type of act that he might perform or has performed.

The sentence "They ought not to go around killing people" is therefore multiply ambiguous. It can mean that one would not expect them to do so (the "ought" of expectation), that it is not in their interest to do so (the "ought" of rationality), that it is a bad thing that they do so (the normative "ought to be"), or that they are wrong to do so (the moral "ought to do"). For the most

---

[2] Thomas Nagel has observed that often, when we use the evaluative "ought to be" to say that something ought to be the case, we imply that someone ought to do something or ought to have done something about it. To take his example, we would not say that a certain hurricane ought not to have killed fifty people just on the ground that it was a terrible thing that the hurricane did so; but we might say this if we had in mind that the deaths from the hurricane would not have occurred except for the absence of safety or evacuation procedures which the authorities ought to have provided.

part I am here concerned only with the last of these interpretations.

The word "should" behaves very much like "ought to." There is a "should" of expectation ("They should be here soon"), a "should" of rationality ("He should go in by the back door"), a normative "should be" ("They shouldn't go around killing people like that"), and the moral "should do" ("You should keep that promise"). I am of course concerned mainly with the last sense of "should."

"Right" and "wrong" also have multiple uses; I will not try to say what all of them are. But I do want to distinguish using the word "wrong" to say that a particular situation or action is wrong from using the word to say that it is wrong *of someone* to do something. In the former case, the word "wrong" is used to assess an act or situation. In the latter case it is used to describe a relation between an agent and an act. Only the latter sort of judgment is an inner judgment. Although we would not say concerning the contented employee of Murder, Incorporated mentioned earlier that it was wrong *of him* to kill Ortcutt, we could say that *his action* was wrong and we could say that it is wrong that there is so much killing.

To take another example, it sounds odd to say that Hitler should not have ordered the extermination of the Jews, that it was wrong of him to have done so. That sounds somehow "too weak" a thing to say. Instead we want to say that Hitler was an evil man. Yet we can properly say, "Hitler ought not to have ordered the extermination of the Jews," if what we mean is that it ought never to have happened; and we can say without oddity that what Hitler did was wrong. Oddity attends only the inner judgment that Hitler was wrong to have acted in that way. That is what sounds "too weak."

It is worth noting that the inner judgments sound too weak not because of the enormity of what Hitler did but because we suppose that in acting as he did he shows that he could not have been susceptible to the moral considerations on the basis of which we make our judgment. He is in the relevant sense beyond the pale and we therefore cannot make inner judgments about him. To see that this is so, consider, say, Stalin, another mass-murderer.

7

We can perhaps imagine someone taking a sympathetic view of Stalin. In such a view, Stalin realized that the course he was going to pursue would mean the murder of millions of people and he dreaded such a prospect; however, the alternative seemed to offer an even greater disaster—so, reluctantly and with great anguish, he went ahead. In relation to such a view of Stalin, inner judgments about Stalin are not as odd as similar judgments about Hitler. For we might easily continue the story by saying that, despite what he hoped to gain, Stalin should not have undertaken the course he did, that it was wrong of him to have done so. What makes inner judgments about Hitler odd, "too weak," is not that the acts judged seem too terrible for the words used but rather that the agent judged seems beyond the pale—in other words beyond the motivational reach of the relevant moral considerations.

Of course, I do not want to deny that for various reasons a speaker might pretend that an agent is or is not susceptible to certain moral considerations. For example, a speaker may for rhetorical or political reasons wish to suggest that someone is beyond the pale, that he should not be listened to, that he can be treated as an enemy. On the other hand, a speaker may pretend that someone is susceptible to certain moral considerations in an effort to make that person or others susceptible to those considerations. Inner judgments about one's children sometimes have this function. So do inner judgments made in political speeches that aim at restoring a lapsed sense of morality in government.

II. THE LOGICAL FORM OF INNER JUDGMENTS

Inner judgments have two important characteristics. First, they imply that the agent has reasons to do something. Second, the speaker in some sense endorses these reasons and supposes that the audience also endorses them. Other moral judgments about an agent, on the other hand, do not have such implications; they do not imply that the agent has reasons for acting that are endorsed by the speaker.

If someone S says that A (morally) ought to do D, S implies that A has reasons to do D and S endorses those reasons—whereas

if *S* says that *B* was evil in what *B* did, *S* does not imply that the reasons *S* would endorse for not doing what *B* did were reasons for *B* not to do that thing; in fact, *S* implies that they were not reasons for *B*.

Let us examine this more closely. If *S* says that (morally) *A* ought to do *D*, *S* implies that *A* has reasons to do *D* which *S* endorses. I shall assume that such reasons would have to have their source in goals, desires, or intentions that *S* takes *A* to have and that *S* approves of *A*'s having because *S* shares those goals, desires, or intentions. So, if *S* says that (morally) *A* ought to do *D*, there are certain motivational attitudes *M* which *S* assumes are shared by *S*, *A*, and *S*'s audience.

Now, in supposing that reasons for action must have their source in goals, desires, or intentions, I am assuming something like an Aristotelian or Humean account of these matters, as opposed, for example, to a Kantian approach which sees a possible source of motivation in reason itself.[3] I must defer a full-scale discussion of the issue to another occasion. Here I simply assume that the Kantian approach is wrong. In particular, I assume that there might be no reasons at all for a being from outer space to avoid harm to us; that, for Hitler, there might have been no reason at all not to order the extermination of the Jews; that the contented employee of Murder, Incorporated might have no reason at all not to kill Ortcutt; that the cannibals might have no reason not to eat their captive. In other words, I assume that the possession of rationality is not sufficient to provide a source for relevant reasons, that certain desires, goals, or intentions are also necessary. Those who accept this assumption will, I think, find that they distinguish inner moral judgments from other moral judgments in the way that I have indicated.

Ultimately, I want to argue that the shared motivational attitudes *M* are intentions to keep an agreement (supposing that others similarly intend). For I want to argue that inner moral judgments are made relative to such an agreement. That is, I want to argue that, when *S* makes the inner judgment that *A*

---

[3] For the latter approach, see Thomas Nagel, *The Possibility of Altruism* (Oxford, 1970).

ought to do D, S assumes that A intends to act in accordance
with an agreement which S and S's audience also intend to
observe. In other words, I want to argue that the source of the
reasons for doing D which S ascribes to A is A's sincere intention
to observe a certain agreement. I have not yet argued for the
stronger thesis, however. I have argued only that S makes his
judgment relative to *some* motivational attitudes M which S
assumes are shared by S, A, and S's audience.

Formulating this as a logical thesis, I want to treat the moral
"ought" as a four-place predicate (or "operator"), "Ought (A,
D, C, M)," which relates an agent A, a type of act D, considera-
tions C, and motivating attitudes M. The relativity to conside-
rations C can be brought out by considering what are sometimes
called statements of prima–facie obligation, "Considering that
you promised, you ought to go to the board meeting, but consi-
dering that you are the sole surviving relative, you ought to go to
the funeral; all things considered, it is not clear what you ought
to do."[4] The claim that there is *this* relativity, to considerations,
is not, of course, what makes my thesis a version of moral rela-
tivism, since any theory must acknowledge relativity to conside-
rations. The relativity to considerations does, however, provide
a model for a coherent interpretation of moral relativism as a
similar kind of relativity.

It is not as easy to exhibit the relativity to motivating attitudes
as it is to exhibit the relativity to considerations, since normally
a speaker who makes a moral "ought" judgment intends the
relevant motivating attitudes to be ones that the speaker shares
with the agent and the audience, and normally it will be obvious
what attitudes these are. But sometimes a speaker does invoke
different attitudes by invoking a morality the speaker does not
share. Someone may say, for example, "As a Christian, you
ought to turn the other cheek; I, however, propose to strike
back." A spy who has been found out by a friend might say, "As
a citizen, you ought to turn me in, but I hope that you will not."
In these and similar cases a speaker makes a moral "ought"

---

4 See Donald Davidson, "Weakness of Will," in Joel Feinberg (ed.), *Moral
Concepts* (Oxford, 1969).

judgment that is explicitly relative to motivating attitudes that the speaker does not share.

In order to be somewhat more precise, then, my thesis is this. "Ought (A, D, C, M)" means roughly that, given that A has motivating attitudes M and given C, D is the course of action for A that is supported by the best reasons. In judgments using this sense of "ought," C and M are often not explicity mentioned by are indicated by the context of utterance. Normally, when that happens, C will be "all things considered" and M will be attitudes that are shared by the speaker and audience.

I mentioned that inner judgments have two characteristics. First, they imply that the agent has reasons to do something that are capable of motivating the agent. Second, the speaker endorses those reasons and supposes that the audience does too. Now, any "Ought (A, D, C, M)" judgment has the first of these characteristics, but as we have just seen a judgment of this sort will not necessarily have the second characteristic if made with explicit reference to motivating attitudes not shared by the speaker. If reference is made either implicitly or explicitly (for example, through the use of the adverb "morally") to attitudes that are shared by the speaker and audience, the resulting judgment has both characteristics and is an inner judgment. If reference is made to attitudes that are not shared by the speaker, the resulting judgment is not an inner judgment and does not represent a full-fledged moral judgment on the part of the speaker. In such a case we have an example of what has been called an inverted-commas use of "ought."[5]

III. Moral Bargaining

I have argued that moral "ought" judgments are relational, "Ought (A, D, C, M)," where M represents certain motivating attitudes. I now want to argue that the attitudes M derive from an agreement. That is, they are intentions to adhere to a particular agreement on the understanding that others also intend to

---

[5] R. M. Hare, *The Language of Morals* (Oxford, 1952), pp. 164-168.

do so. Really, it might be better for me to say that I put this forward as a hypothesis, since I cannot pretend to be able to prove that it is true. I will argue, however, that this hypothesis accounts for an otherwise puzzling aspect of our moral views that, as far as I know, there is no other way to account for.

I will use the word "intention" in a somewhat extended sense to cover certain dispositions or habits. Someone may habitually act in accordance with the relevant understanding and therefore may be disposed to act in that way without having any more or less conscious intention. In such a case it may sound odd to say that he *intends* to act in accordance with the moral understanding. Nevertheless, for present purposes I will count that as his having the relevant intention in a dispositional sense.

I now want to consider the following puzzle about our moral views, a puzzle that has figured in recent philosophical discussion of issues such as abortion. It has been observed that most of us assign greater weight to the duty not to harm others than to the duty to help others. For example, most of us believe that a doctor ought not to save five of his patients who would otherwise die by cutting up a sixth patient and distributing his healthy organs where needed to the others, even though we do think that the doctor has a duty to try to help as many of his patients as he can. For we also think that he has a stronger duty to try not to harm any of his patients (or anyone else) even if by so doing he could help five others.[6]

This aspect of our moral views can seem very puzzling, especially if one supposes that moral feelings derive from sympathy and concern for others. But the hypothesis that morality derives from an agreement among people of varying powers and resources provides a plausible explanation. The rich, the poor, the strong, and the weak would all benefit if all were to try to avoid harming one another. So everyone could agree to that arrangement. But the rich and the strong would not benefit from an arrangement whereby everyone would try to do as much as possible to help those in need. The poor and weak would get all of the benefit of

---

[6] Philippa Foot, "Abortion and the Doctrine of Double Effect," in James Rachels (ed.), *Moral Problems* (New York, 1971).

this latter arrangement. Since the rich and the strong could foresee that they would be required to do most of the helping and that they would receive little in return, they would be reluctant to agree to a strong principle of mutual aid. A compromise would be likely and a weaker principle would probably be accepted. In other words, although everyone could agree to a strong principle concerning the avoidance of harm, it would not be true that everyone would favor an equally strong principle of mutual aid. It is likely that only a weaker principle of the latter sort would gain general acceptance. So the hypothesis that morality derives from an understanding among people of different powers and resources can explain (and, according to me, does explain) why in our morality avoiding harm to others is taken to be more important than helping those who need help.

By the way, I am here only trying to *explain* an aspect of our moral views. I am not therefore *endorsing* that aspect. And I defer until later a relativistic account of the way in which aspects of our moral view can be criticized "from within."

Now we need not suppose that the agreement or understanding in question is explicit. It is enough if various members of society knowingly reach an agreement in intentions—each intending to act in certain ways on the understanding that the others have similar intentions. Such an implicit agreement is reached through a process of mutual adjustment and implicit bargaining.

Indeed, it is essential to the proposed explanation of this aspect of our moral views to suppose that the relevant moral understanding is thus the result of *bargaining*. It is necessary to suppose that, in order to further our interests, we form certain conditional intentions, hoping that others will do the same. The others, who have different interests, will form somewhat different conditional intentions. After implicit bargaining, some sort of compromise is reached.

Seeing morality in this way as a compromise based on implicit bargaining helps to explain why our morality takes it to be worse to harm someone than to refuse to help someone. The explanation requires that we view our morality as an implicit agreement about what to do. This sort of explanation could not be given if we were to suppose, say, that our morality represented an agree-

ment only about the facts (naturalism). Nor is it enough simply
to suppose that our morality represents an agreement in attitude,
if we forget that such agreement can be reached, not only by way
of such principles as are mentioned, for example, in Hare's
"logic of imperatives,"[7] but also through bargaining. According
to Hare, to accept a general moral principle is to intend to do
something.[8] If we add to his theory that the relevant intentions
can be reached through implicit bargaining, the resulting theory
begins to look like the one that I am defending.

Many aspects of our moral views can be given a utilitarian
explanation. We could account for these aspects, using the logical
analysis I presented in the previous section of this paper, by
supposing that the relevant "ought" judgments presuppose
shared attitudes of sympathy and benevolence. We can equally
well explain them by supposing that considerations of utility
have influenced our implicit agreements, so that the appeal is to
a shared intention to adhere to those agreements. Any aspect of
morality that is susceptible of a utilitarian explanation can also
be explained by an implicit agreement, but not conversely.
There are aspects of our moral views that seem to be explicable
only in the second way, on the assumption that morality derives
from an agreement. One example, already cited, is the distinction
we make between harming and not helping. Another is our
feeling that each person has an inalienable right of self-defense
and self-preservation. Philosophers have not been able to come
up with a really satisfactory utilitarian justification of such a
right, but it is easily intelligible on our present hypothesis, as
Hobbes observed many years ago. You cannot, except in very
special circumstances, rationally form the intention not to try to
preserve your life if it should ever be threatened, say, by society
or the state, since you know that you cannot now control what
you would do in such a situation. No matter what you now
decided to do, when the time came, you would ignore your prior
decision and try to save your life. Since you cannot now intend
to do something later which you now know that you would not

[7] R. M. Hare, *op. cit.* and *Freedom and Reason* (Oxford, 1963).
[8] *The Language of Morals*, pp. 18-20, 168-169.

do, you cannot now intend to keep an agreement not to preserve your life if it is threatened by others in your society.[9]

This concludes the positive side of my argument that what I have called inner moral judgments are made in relation to an implicit agreement. I now want to argue that this theory avoids difficulties traditionally associated with implicit agreement theories of morality.

IV. Objections and Replies

One traditional difficulty for implicit agreement theories concerns what motivates us to do what we have agreed to do. It will, obviously, not be enough to say that we have implicitly agreed to keep agreements, since the issue would then be why we keep *that* agreement. And this suggests an objection to implicit agreement theories. But the apparent force of the objection derives entirely from taking an agreement to be a kind of ritual. To agree in the relevant sense is not just to say something; it is to intend to do something—namely, to intend to carry out one's part of the agreement on the condition that others do their parts. If we agree in this sense to do something, we intend to do it and intending to do it is already to be motivated to do it. So there is no problem as to why we are motivated to keep our agreements in this sense.

We do believe that in general you ought not to pretend to agree in this sense in order to trick someone else into agreeing. But that suggests no objection to the present view. All that it indicates is that *our* moral understanding contains or implies an agreement to be open and honest with others. If it is supposed that this leaves a problem about someone who has not accepted our agreement—"What reason does *he* have not to pretend to accept our agreement so that he can then trick others into agreeing to various things?"—the answer is that such a person may or may not have such a reason. If someone does not already accept something of our morality it may or may not be possible to find reasons why he should.

---

[9] Cf. Thomas Hobbes, *Leviathan* (Oxford, 1957, *inter alia*), Pt. I, Ch. 14, "Of the First and Second Natural Laws, And of Contracts."

A second traditional objection to implicit agreement theories is that there is not a perfect correlation between what is generally believed to be morally right and what actually is morally right. Not everything generally agreed on is right and sometimes courses of action are right that would not be generally agreed to be right. But this is no objection to my thesis. My thesis is not that the implicit agreement from which a morality derives is an agreement in moral judgment; the thesis is rather that moral judgments make reference to and are made in relation to an agreement in intentions. Given that a group of people have agreed in this sense, there can still be disputes as to what the agreement implies for various situations. In my view, many moral disputes are of this sort. They presuppose a basic agreement and they concern what implications that agreement has for particular cases.

There can also be various things wrong with the agreement that a group of people reach, even from the point of view of that agreement, just as there can be defects in an individual's plan of action even from the point of view of that plan. Given what is known about the situation, a plan or agreement can in various ways be inconsistent, incoherent, or self-defeating. In my view, certain moral disputes are concerned with internal defects of the basic moral understanding of a group, and what changes should be made from the perspective of that understanding itself. This is another way in which moral disputes make sense with reference to and in relation to an underlying agreement.

Another objection to implicit agreement theories is that not all agreements are morally binding—for example, those made under complusion or from a position of unfair disadvantage, which may seem to indicate that there are moral principles prior to those that derive from an implicit agreement. But, again, the force of the objection derives from an equivocation concerning what an agreement is. The principle that compelled agreements do not obligate concerns agreement in the sense of a certain sort of ritual indicating that one agrees. My thesis concerns a kind of agreement in intentions. The principle about compelled agreements is part of, or is implied by, our agreement in intentions.

According to me it is only with reference to some such agreement in intentions that a principle of this sort makes sense.

Now it may be true our moral agreement in intentions also implies that it is wrong to compel people who are in a greatly inferior position to accept an agreement in intentions that they would not otherwise accept, and it may even be true that there is in our society at least one class of people in an inferior position who have been compelled thus to settle for accepting a basic moral understanding, aspects of which they would not have accepted had they not been in such an inferior position. In that case there would be an incoherence in our basic moral understanding and various suggestions might be made concerning the ways in which this understanding should be modified. But this moral critique of the understanding can proceed from that understanding itself rather than from "prior" moral principles.

In order to fix ideas, let us consider a society in which there is a well-established and long-standing tradition of hereditary slavery. Let us suppose that everyone accepts this institution, including the slaves. Everyone treats it as in the nature of things that there should be such slavery. Furthermore, let us suppose that there are also aspects of the basic moral agreement which speak against slavery. That is, these aspects together with certain facts about the situation imply that people should not own slaves and that slaves have no obligation to acquiesce in their condition. In such a case, the moral understanding would be defective, although its defectiveness would presumably be hidden in one or another manner, perhaps by means of a myth that slaves are physically and mentally subhuman in a way that makes appropriate the sort of treatment elsewhere reserved for beasts of burden. If this myth were to be exposed, the members of the society would then be faced with an obvious incoherence in their basic moral agreement and might come eventually to modify their agreement so as to eliminate its acceptance of slavery.

In such a case, even relative to the old agreement it might be true that slave owners ought to free their slaves, that slaves need not obey their masters, and that people ought to work to eliminate slavery. For the course supported by the best reasons, given that one starts out with the intention of adhering to a particular

agreement, may be that one should stop intending to adhere to certain aspects of that agreement and should try to get others to do the same.

We can also (perhaps—but see below) envision a second society with hereditary slavery whose agreement has no aspects that speak against slavery. In that case, even if the facts of the situation were fully appreciated, no incoherence would appear in the basic moral understanding of the society and it would not be true in relation to that understanding that slave owners ought to free their slaves, that slaves need not obey their masters, and so forth. There might nevertheless come a time when there were reasons of a different sort to modify the basic understanding, either because of an external threat from societies opposed to slavery or because of an internal threat of rebellion by the slaves.

Now it is easier for us to make what I have called inner moral judgments about slave owners in the first society than in the second. For we can with reference to members of the first society invoke principles that they share with us and, with reference to those principles, we can say of them that they ought not to have kept slaves and that they were immoral to have done so. This sort of inner judgment becomes increasingly inappropriate, however, the more distant they are from us and the less easy it is for us to think of our moral understanding as continuous with and perhaps a later development of theirs. Furthermore, it seems appropriate to make only non-inner judgments of the slave owners in the second society. We can say that the second society is unfair and unjust, that the slavery that exists is wrong, that it ought not to exist. But it would be inappropriate in this case to say that it was morally wrong of the slave owners to own slaves. The relevant aspects of our moral understanding, which we would invoke in moral judgments about them, are not aspects of the moral understanding that exists in the second society. (I will come back to the question of slavery below.)

Let me turn now to another objection to implicit agreement theories, an objection which challenges the idea that there is an agreement of the relevant sort. For, if we have agreed, when did we do it? Does anyone really remember having agreed? How did we indicate our agreement? What about those who do not

want to agree? How do they indicate that they do not agree and what are the consequences of their not agreeing? Reflection on these and similar questions can make the hypothesis of implicit agreement seem too weak a basis on which to found morality.

But once again there is equivocation about agreements. The objection treats the thesis as the claim that morality is based on some sort of ritual rather than an agreement in intentions. But, as I have said, there is an agreement in the relevant sense when each of a number of people has an intention on the assumption that others have the same intention. In this sense of "agreement," there is no given moment at which one agrees, since one continues to agree in this sense as long as one continues to have the relevant intentions. Someone refuses to agree to the extent that he or she does not share these intentions. Those who do not agree are outside the agreement; in extreme cases they are outlaws or enemies. It does not follow, however, that there are no constraints on how those who agree may act toward those who do not, since for various reasons the agreement itself may contain provisions for dealing with outlaws and enemies.

This brings me to one last objection, which derives from the difficulty people have in trying to give an explicit and systematic account of their moral views. If one actually agrees to something, why is it so hard to say what one has agreed? In response I can say only that many understandings appear to be of this sort. It is often possible to recognize what is in accordance with the understanding and what would violate it without being able to specify the understanding in any general way. Consider, for example, the understanding that exists among the members of a team of acrobats or a symphony orchestra.

Another reason why it is so difficult to give a precise and systematic specification of any actual moral understanding is that such an understanding will not in general be constituted by absolute rules but will take a vaguer form, specifying goals and areas of responsibility. For example, the agreement may indicate that one is to show respect for others by trying where possible to avoid actions that will harm them or interfere with what they are doing; it may indicate the duties and responsibilities of various members of the family, who is to be responsible for bringing up the

children, and so forth. Often what will be important will be not so
much exactly what actions are done as how willing participants
are to do their parts and what attitudes they have—for example,
whether they give sufficient weight to the interests of others.

The vague nature of moral understandings is to some extent
alleviated in practice. One learns what can and cannot be done
in various situations. Expectations are adjusted to other expec-
tations. But moral disputes arise nonetheless. Such disputes may
concern what the basic moral agreement implies for particular
situations; and, if so, that can happen either because of disputes
over the facts or because of a difference in basic understanding.
Moral disputes may also arise concerning whether or not changes
should be made in the basic agreement. Racial and sexual issues
seem often to be of this second sort; but there is no clear line
between the two kinds of dispute. When the implications of an
agreement for a particular situation are considered, one possible
outcome is that it becomes clear that the agreement should be
modified.

Moral reasoning is a form of practical reasoning. One begins
with certain beliefs and intentions, including intentions that are
part of one's acceptance of the moral understanding in a given
group. In reasoning, one modifies one's intentions, often by
forming new intentions, sometimes by giving up old ones, so
that one's plans become more rational and coherent—or, rather,
one seeks to make all of one's attitudes coherent with each other.

The relevant sort of coherence is not simply consistency. It is
something very like the explanatory coherence which is so im-
portant in theoretical reasoning. Coherence involves generality
and lack of arbitrariness. Consider our feelings about cruelty to
animals. Obviously these do not derive from an agreement that
has been reached with animals. Instead it is a matter of coherence.
There is a prima-facie arbitrariness and lack of generality in a
plan that involves avoiding cruelty to people but not to animals.

On the other hand, coherence in this sense is not the only
relevant factor in practical reasoning. Another is conservatism
or inertia. A third is an interest in satisfying basic desires or needs.
One tries to make the least change that will best satisfy one's
desires while maximizing the overall coherence of one's atti-

tudes. Coherence by itself is not an overwhelming force. That is why our attitudes towards animals are weak and wavering, allowing us to use them in ways we would not use people.

Considerer again the second hereditary slave society mentioned above. This society was to be one in which no aspects of the moral understanding shared by the masters spoke against slavery. In fact that is unlikely, since there is *some* arbitrariness in the idea that people are to be treated in different ways depending on whether they are born slave or free. Coherence of attitude will no doubt speak at least a little against the system of slavery. The point is that the factors of conservatism and desire might speak more strongly in favor of the *status quo*, so that, all things considered, the slave owners might have no reason to change their understanding.

One thing that distinguishes slaves from animals is that slaves can organize and threaten revolt, whereas animals cannot. Slaves can see to it that both coherence and desire oppose conservatism, so that it becomes rational for the slave owners to arrive at a new, broader, more coherent understanding, one which includes the slaves.

It should be noted that coherence of attitude provides a constant pressure to widen the consensus and eliminate arbitrary distinctions. In this connection it is useful to recall ancient attitudes toward foreigners, and the ways people used to think about "savages," "natives," and "Indians." Also, recall that infanticide used to be considered as acceptable as we consider abortion to be. There has been a change here in our moral attitudes, prompted, I suggest, largely by considerations of coherence of attitude.

Finally, I would like to say a few brief words about the limiting case of group morality, when the group has only one member; then, as it were, a person comes to an understanding with himself. In my view, a person can make inner judgments in relation to such an individual morality only about himself. A familiar form of pacifism is of this sort. Certain pacifists judge that it would be wrong of them to participate in killing, although they are not willing to make a similar judgment about others. Observe that such a pacifist is unwilling only to make *inner* moral judgments about others. Although he is unwilling to judge that those who do

GILBERT HARMAN

participate are wrong to do so, he is perfectly willing to say that
it is a bad thing that they participate. There are of course many
other examples of individual morality in this sense, when a person
imposes standards on himself that he does not apply to others.
The existence of such examples is further confirmation of the
relativist thesis that I have presented.

My conclusion is that relativism can be formulated as an
intelligible thesis, the thesis that morality derives from an implicit
agreement and that moral judgments are in a logical sense made
in relation to such an agreement. Such a theory helps to explain
otherwise puzzling aspects of our own moral views, in particular
why we think that it is more important to avoid harm to others
than to help others. The theory is also partially confirmed by
what is, as far as I can tell, a previously unnoticed distinction
between inner and non-inner moral judgments. Furthermore,
traditional objections to implicit agreement theories can be met.[10]

GILBERT HARMAN

Princeton University

[10] Many people have given me good advice about the subjects discussed in
this paper, which derives from a larger study of practical reasoning and moral-
ity. I am particularly indebted to Donald Davidson, Stephen Schiffer, William
Alston, Frederick Schick, Thomas Nagel, Walter Kaufmann, Peter Singer,
Robert Audi, and the editors of the *Philosophical Review*.