The Limits of Objectivity

THOMAS NAGEL

THE TANNER LECTURE ON HUMAN VALUES

Delivered at
Brasenose College, Oxford University

May 4, 11, and 18, 1979

THOMAS NAGEL is Professor of Philosophy at Princeton University. He was educated at Cornell, Oxford, and Harvard, and taught from 1963 to 1966 at the University of California, Berkeley, before moving to Princeton. He has been a Fellow of the Guggenheim and National Science Foundations. Professor Nagel is the author of *The Possibility of Altruism* (1970) and *Mortal Questions* (1979), and associate editor of *Philosophy and Public Affairs* .

I. THE MIND

1. These lectures are about objectivity and its limits. In the second and third lectures I shall be concerned with normative questions; I shall defend the objectivity of ethics, and try to explain what it means. But today I am going to say something about the problem of objectivity as it occurs in metaphysics, especially in the philosophy of mind. I do this because the problem has a similar form in the two areas, and because ideas arising from metaphysics influence our views of what must be done to discover objectivity in ethics. I hope therefore not only to say something about subjectivity and objectivity in the philosophy of mind, but also to set the stage for an account of what it would be for ethics to be objective.

2. As an aid to comprehension, let me begin by asserting without argument what I hope to show by examination of particular cases.

Objectivity is a method of understanding. It is beliefs and knowledge that are objective in the primary sense. Only derivatively do we call objective the truths that can be understood in this way.

To acquire a more objective understanding of some aspect of the world, we step back from our view of it and form a new conception which has that view and its relation to the world as its object. In other words, we place ourselves in the world that is to be understood. The old view then comes to be regarded as an appearance, more subjective than the new view, and correctable or confirmable by reference to it. The process can be repeated, yielding a still more objective conception.

But it will not always yield a result, and sometimes it will be thought to yield a result when it really doesn't: then, as Nietzsche warned, one will get a false objectification of an aspect of reality that cannot be better understood from a more objective standpoint. So although there is a connection between objectivity and reality — only the supposition that we and our appearances are parts of a larger reality makes it reasonable to seek understanding by stepping back from the appearances in this way — still not all reality is objective, for not everything is better understood the more objectively it is viewed. Appearance and perspective are essential parts of what there is, and in some respects they are best understood from a less detached standpoint. Both in ethics and in metaphysics, I believe, realism underlies the claims of objectivity and detachment, but it supports them only up to a point.

3. The question I want to discuss now is whether there is a sense in which the mind and the self are parts of objective reality. Eventually I shall take up the question of what it is for a particular person to be *me* (or *you*). But first I am going to talk about the objective status of mental phenomena in general.

This question is in the background of the mind–body problem, for the mind–body problem arises because certain features of subjective experience resist accommodation by one very important conception of objectivity. I am not going to offer a solution to the mind–body problem here. But I believe that no progress can be made with it unless we understand this conception and examine its claims with care.

For convenience I shall refer to it as the *physical* conception of objectivity. It is not the same thing as our idea of what physical reality is actually like, but it has developed as part of our method of arriving at a truer understanding of the physical world, a world that is presented to us initially but somewhat inaccurately through sensory perception.

The development goes in stages, each of which gives a more objective picture than the one before. The first step is to see that our perceptions are caused by the action of things on us, through their effects on our bodies, which are themselves parts of the physical world. The next step is to realize that since the same physical properties that cause perceptions in us through our bodies also produce different effects on other physical things and can exist without causing any perceptions at all, their true nature must be detachable from their perceptual appearance and need not resemble it. The third step is to try to form a conception of that true nature independent of its appearance either to us or to other types of perceivers. This means not only not thinking of the physical world from your own particular point of view, but not thinking of it from a more general human perceptual point of view either: not thinking of how it looks, feels, smells, tastes, or sounds. The secondary qualities then drop away, and the primary qualities are thought of structurally.

This has turned out to be an extremely fruitful strategy. The understanding of the physical world has been expanded enormously with the aid of theories and explanations that use concepts not tied to a specifically human perceptual viewpoint. Our senses provide the evidence from which we start, but the detached character of this understanding is such that we could possess it even if we had none of our present senses, so long as we were rational and could understand the mathematical and formal properties of the objective conception of the physical world. We might even in a sense share an understanding of physics with other creatures to whom things appeared quite different, perceptually — so long as they too were rational and numerate.

The world described by this objective conception is not just centerless, it is also in a sense featureless. While the things in it have properties, none of these properties are perceptual aspects. All of those have been relegated to the mind, a yet-to-be-examined domain. The physical world as it is supposed to be in itself con-

tains no points of view and nothing that can appear only to a particular point of view. Whatever it contains can be apprehended by a general rational consciousness that gets its information through whichever perceptual point of view it happens to view the world from.[1]

4. Powerful as it has proven to be, this bleached-out physical conception of objectivity encounters difficulties if it is put forward as the method for seeking a complete understanding of reality. For the process began when we noticed that how things appear to us depends on the interaction of our bodies with the rest of the world. But this leaves us with no account of the perceptions and specific viewpoints which were left behind as irrelevant to physics but which seem to exist nonetheless, along with those of other creatures. Not to mention the mental activity of forming an objective conception of the physical world, which seems not itself capable of physical analysis.

Faced with these facts one might think the only conceivable conclusion would be that there is more to reality than what can be accommodated by the physical conception of objectivity. But to remarkable numbers of people this has not been obvious. The physical has been so irresistibly attractive, and has so dominated ideas of what there is, that attempts have been made to beat everything into its shape and deny the reality of anything that cannot be so reduced. As a result the philosophy of mind is populated with extremely implausible positions.

I think part of the explanation of this modern weakness for reduction is that a less impoverished and reductive idea of objectivity has not been available, to fill out the project of constructing an objective picture of the world. The objectivity of physics was viable: it continued to yield progressively more understanding

---

[1] Bernard Williams gives an excellent account of this idea in his recent book, *Descartes: The Project of Pure Inquiry* (Harmondsworth, Middlesex: Penguin Books, 1978). He calls it the *absolute* conception of reality.

through successive applications to those properties of the physical world that earlier applications had discovered.

It is true that recent developments in physics have led some to believe that it may after all be incapable of providing a conception of what is really there, independently of observation. But I do not wish to argue that since the idea of objective reality has to be abandoned because of quantum theory anyway, we might as well go the whole hog and admit the subjectivity of the mental. Even if, as some physicists think, quantum theory cannot be interpreted in a way that permits the phenomena to be described without reference to an observer, the heliminable observer need not be a member of any particular species, like the human, to whom things look and feel in highly characteristic ways. This does not therefore require that we let in the full range of subjective experience.

The central problem is not whether points of view must be admitted to the account of the *physical* world. Whatever may be the answer to that question, we shall still be faced with an independent problem about the mind. It is the phenomena of consciousness themselves that pose the clearest challenge to the idea that physical objectivity gives the general form of reality. In response I do not want to abandon the idea of objectivity entirely but rather to suggest that the physical is not its only possible interpretation.

5. Eventually I shall argue that the claims of even an expanded objectivity should not be exaggerated. But first I want to explore the possibility of arriving at an objective concept of mind. The reason for wanting such a thing is that we assume some connection, even if not a very tight one, between what is real and what can be objectively understood. We assume in particular that we ourselves, minds included, are parts of the world as it is in itself, and not *just* parts of the world as it appears to us: though we are of course *also* parts of that phenomenal world. And if we are

parts of the world as it is in itself, then we would hope to be able to acquire some conception of ourselves that is not just the conception from within: a conception of ourselves from without, as *contained in* the world.

It is not obvious that this is possible, but it is natural enough to consider whether it might be. And to find out, we must ask ourselves whether it is possible to form a conception of our own minds from an objective standpoint. Here it is essential, as it is in regard to other matters, not to identify objectivity with the *physical conception* of objectivity. We have to think of objectivity as something general enough to admit of different interpretations for different subjects of inquiry.

The general idea of objectivity that we must use to think about a single world containing both mental and physical phenomena, is the idea of the world as it is, rather than as it appears to any particular viewpoint within it. Even if such a conception works very differently with respect to minds from the way it works with respect to matter, it still has to provide a way of thinking about what the world contains in detachment from any particular point of view within that world. The results will be understandable to individuals who occupy various points of view only if they can think about the world in detachment from their particular perspectives on it.

Our capacity for such detachment, indeed our appetite for it, is one of our most important and creative characteristics. It leads to false objective conceptions of mind and other things if the supply of interpretations of objectivity is too meagre. But we may be able to remedy this if we try to develop an interpretation to suit the subject matter instead of trying to understand the mind and its attributes by means of a conception of objectivity developed to account for a completely different set of things. The question then is whether, viewing the world in detachment from our particular perspective on it —i.e., viewing it as a place that *contains* us —we can form an objective conception that *includes* points of

view, our own and those of others, and that does not misrepresent them as aspects of objective *physical* reality.

There are many points of view in the world, of many different kinds, and we are familiar with only a few. To understand them all from no particular point of view would seem to require a process of self-transcendence different from that which occurs when we investigate the external world. Perhaps no conception of objectivity adequate to the mind can exist, in which case we shall have to choose between abandoning the assumption that everything real has an objective character and abandoning the assumption that the mind is real. But we are not faced with this dilemma simply by the failure of the *physical* conception of objectivity to apply to the mental. It is clear that an objective conception of mental phenomena cannot, like that of physical phenomena, be based on abstraction from the specific form of our *external* perception of them. So we must ask instead whether there can be an understanding of them independent of the specific point of view to which they appear, which nevertheless keeps their perspectival character.

What I want to do is to explain what a natural objective understanding of the mind along these lines would be. I believe it has its beginnings in the ordinary concept of mind, but that it can be developed beyond this. The question is, how far beyond?

In my view, quite far. I believe we can include ourselves, experiences and all, in a world conceivable not from a specifically human point of view, and that we can do this without reducing the mental to the physical. But I also believe that any such conception will necessarily be incomplete. And this means that the pursuit of an objective conception of reality comes up against limits that are not merely practical, limits that could not be overcome by any objective intelligence, however powerful. But finally, I shall claim that this is no cause for philosophical alarm, because there is no reason to assume that the world as it is in itself must be objectively comprehensible. It is natural for us to want to bring

our capacity for detached, objective understanding as much into alignment with reality as we can, but it should not surprise us if objectivity is essentially *incomplete.*

6. The fundamental problem of how the mind can be objectively understood appears in philosophy independently of the ambition to form a complete scientific conception of the world. It appears as the problem of other minds. Each of us is the subject of various experiences, and to understand that there are other people in the world as well, one must be able to conceive of experiences of which one is not the subject: experiences that are not *present* to oneself. To do this it is necessary to have a general conception of subjects of experience and to place oneself under it as an instance. It will not do simply to extend the idea of what is immediately felt into other people's bodies, for as Wittgenstein observed, that will only give you an idea of feeling things *in their bodies,* not of *their* feeling things.

Though we all grow up with the required general conception that allows us to believe in genuinely other minds, it has been philosophically very problematic, and there has been much difference of opinion over how it works. Some philosophers have been attracted to analyses in behavioral, causal, or functional terms which are objective in the sense in which physics is objective. That is because, though nontheoretical, the ordinary concept of mind is evidently a conception of how things are, and it is assumed that physical objectivity provides the general form of understanding how things are.

Others, seeing that the physical idea of objectivity cannot be applied to the mental, have been left with an insoluble problem of solipsism: the inability to make sense of the idea of real minds other than one's own. Solipsism seems to me to represent a higher level of insight than reductionism, for it does not throw away the problem which must be faced and solved if we are to understand what minds are and how the world can contain them.

But both these responses rest on a fundamental mistake. The ordinary concept of mind contains the beginnings of an entirely different way of conceiving objective reality. We do not make sense of the idea of other minds by construing it in a way which becomes unintelligible when we try to apply it to ourselves. We do not abandon the essential factor of a point of view when we conceive of the minds of others: instead we generalize it, and think of ourselves as one point of view among others. The first stage of objectification of the mental is for each of us to be able to grasp the idea of all human perspectives, including his own, without depriving them of their character as perspectives. It is the analogue for minds of a centerless conception of space for physical objects, in which no point has a privileged position.

The beginning of an objective concept of mind is the ability to view one's own experiences from outside, as events in the world. If this is possible, then others can also conceive of those events and one can conceive of the experiences of others, also from outside. Any experience can be thought about and known to have occurred not only from the point of view of its subject but from other points of view, at least if they are sufficiently similar in type to that of the subject.

To think in this way we use not a faculty of external representation, but a general idea of subjective points of view, of which we imagine a particular instance and a particular form. It is this general faculty of sympathetic subjective imagination that takes us on the first step outside of ourselves in the acquisition of an objective concept of mind, and that enables each person to place himself among the contents of the world.

So far the process does not involve any abstraction from the general forms of our experience. We still think of experience in terms of the familiar point of view we share with other humans. All that is involved in the external conception of mind is the imaginative use of this point of view — a use that is partly present in the memory and expectation of one's own experiences.

To represent an experience from outside by imagining it subjectively is the analogue of representing an objective spatial configuration by imagining it visually. One uses ordinary appearance as a medium. What is represented does not resemble the representation in all respects. It is represented in terms of certain general features of subjective experience —subjective universals — some instances of which one is familiar with from one's own experience. But the capacity to form universal concepts in any area enables one not only to represent the present situation from without but to think about other possibilities which one has not experienced and perhaps never will experience directly. So the pretheoretical concept of mind involves a kind of objectivity which permits us to go some way beyond our own experiences and those exactly like them. The difficult step is the next one, the step beyond representation by resemblance.

7. Of course one possibility is that this particular process can go no farther. We can have a concept of mind general enough to allow us to escape solipsism and perhaps even ethnocentrism, but perhaps we cannot transcend the general forms of human experience and the human viewpoint. That viewpoint permits us to conceive of experiences we have not had, because of the flexibility of the human imagination. But it may not allow us to detach the concept of mind from a *human perspective.* If this is so, then there are strict limits to the objectivity of that concept; and that seems to mean that we cannot conceive of ourselves as parts of a world whose reality can be acknowledged from every rational perspective.

This is a drastic conclusion, but I think it follows if we restrict the pursuit of objectivity in this area to the use of subjective imagination about the experiences of others. Even allowing for a certain amount of flexibility, the subjective imagination can reach only so far. And it may seem that there is no other way of conceiving of minds —our own and others' —from outside with-

out losing hold of the fact that they are perspectives or points of view. For if we don't conceive of them from inside, we must be conceiving of them as part of the familiar external world, and that is the old mistake.

So the issue is whether we can form a general concept of experience that extends far beyond our own or anything like it. Or more accurately, whether there can *be* such a concept — for we may be unable to grasp it as we are presumably unable to grasp now concepts of objective physical reality which will be developed five centuries hence. The possibility that there is such a concept is sufficient motive for trying to form it. It is only if we are convinced in advance that the thing makes no sense that we can be justified in setting the limits of objectivity with regard to the mind so close to our own ordinary viewpoint.

8. I believe that in fact we already possess a rudimentary general concept of experience, and that it does not lose all content when we use it to think about cases in which we cannot apply it more specifically.

Consider first, cases where we have strong evidence that experience is present, without either knowing what its character is, or even being in a position to hope ever to reach an understanding of its character that will include the capacity for self-ascription. This is true of at least some of the experiences of all animals that are not very close to us in structure and behavior. In each case there is extensive external evidence of conscious inner life, but only limited application of our own mental concepts — mostly general ones — to describe it.[2]

It is the ordinary prephilosophical concept of experience that leads to this result. We have not simply left it behind and taken off with the *word*. And the extension is not part of a private

---

[2] Skeptics should read Herbert Spencer Jennings' great book, originally published in 1906, *The Behavior of the Lower Organisms* (Bloomington: Indiana University Press, 1976).

language but a natural idea shared by most human beings about what sorts of things occupy the world around them. We are forced, I think, to conclude that all these creatures have specific experiences which cannot be represented by any mental concepts of which we could have first-person understanding. This doesn't mean that we can't think about them in that general way, or perhaps in more detail but without first-person understanding — provided that we continue to regard them as subjective experiences rather than mere behavioral dispositions or functional states.

But it seems to me that we can in principle go farther. We can use the general concepts of experience and mind to speculate about forms of conscious life whose external signs we cannot confidently identify. There is probably a great deal of life in the universe, and we may be in a position to identify only some of its forms, because we would simply be unable to read as behavior the manifestations of creatures sufficiently unlike us. It certainly means something to speculate that there are such creatures, and that they have minds.

These uses of the general concept of mind exemplify a theoretical step that is commonplace elsewhere. We can form the idea of phenomena that we do not know how to detect. Once the conception of a new physical particle is formed, defined in terms of a set of properties, those properties may then allow experiments to be devised which will permit its detection. In this way the progress of physical discovery has long since passed to the formation of physical concepts that can be applied only with sophisticated techniques of observation, and not by means of unaided perception or simple mechanical measurement.

Only an unacceptable verificationist dogmatism would deny the possibility of forming objective concepts that reach beyond our current capacity to apply them. The aim of reaching a conception of the world which does not put us at the center in any way requires the formation of such concepts. We are supported in this aim by a kind of intellectual optimism: the belief that we possess

an open-ended capacity for understanding what we have not yet conceived, and that it can be called into operation by detaching from our present understanding and trying to reach a higher-order view which explains it as part of the world.

It is the same with the mind. To accept the *general* idea of a perspective without limiting it to the forms with which one is familiar, subjectively or otherwise, is the precondition of seeking ways to conceive of particular types of experience that do not depend on the ability either to have those experiences or to imagine them subjectively. It should be possible to investigate in this way the quality-structure of some sense we do not have, for example, by observing creatures who do have it.

But if we could do that, we should also be able to apply the same general idea to ourselves, and thus to analyze our experiences in ways that can be understood without having had such experiences. That would constitute a kind of objective standpoint toward our own minds. To the extent that it could be achieved, we would be able to see ourselves as not merely part of the human world: something we can already do with regard to our bodies. And this would serve a natural human goal: for it is natural to want to reach a general understanding of reality, including ourselves, which does not depend on the fact that we *are* ourselves.

9. In the pursuit of this goal, however, even at its most successful, something will inevitably be lost. If we try to understand experience from an objective viewpoint that is distinct from that of the subject of the experience, then even if we continue to credit its perspectival nature, we will not be able to grasp its most specific qualities unless we can imagine them subjectively. We will not know exactly how scrambled eggs taste to a cockroach even if we develop a very complete objective phenomenology of his sense of taste. When it comes to values, goals, and forms of life the gulf may be even more profound.

Since this is so, no objective conception of the mental world

can include it all. But in that case we must ask what the point is of looking for such a conception. The aim was to place perspectives and their contents in a world seen from no particular point of view. It turns out that some aspects of those perspectives cannot be understood in terms of an objective concept of mind. So with respect to those aspects we are in the same position we were in to begin with. We must say either that they are not real or that not everything real is part of objective reality. Since the first is patently absurd, that leaves the second. But if some aspects of reality can't be captured in an objective conception, why not forget the ambition of capturing as much of it as possible? The world just *isn't* the world as it appears to one highly abstracted point of view. And if one can't have complete objectivity, the goal of capturing as much of reality as one can in an objective net is pointless and unmotivated.

I believe there is an answer to this, and that it is the answer to a very general problem of which this is an instance. Reality is not just objective reality, and any objective conception, in order not to be false, must include an acknowledgment of its own incompleteness. This is an important qualification to the claims of objectivity in other areas as well, and later I shall argue that it has a direct application to ethics. We may try to develop as complete an understanding of values as we can from a neutral standpoint but we will have to acknowledge the existence and validity of some values that cannot be neutrally understood, and even of some values that cannot be either neutrally or sympathetically understood.

But to return to the case under discussion: even if an objective conception of mind were developed, it would have to include the qualification that the exact character of each of the experiential and intentional perspectives with which it deals can be understood only from within or by subjective imagination. A being with total imaginative power could understand it all, but an ordinary being using an objective concept of mind will not.

So we have not given up the idea of the way the world really is, independently of how it appears to us or to any particular occupant of it. We have only given up the idea that this coincides with what can be objectively understood. The way the world is includes appearances, and there is no single point of view from which they can all be fully grasped.

This amounts to the rejection of idealism with regard to the mind. The world is not my world, or our world —not even the mental world is. This is a particularly unequivocal rejection of idealism because it affirms the reality of aspects of the world that cannot be grasped by *any* conception I can possess —not even an objective conception of the kind with which we transcend the domain of initial appearances. Here it can be seen that physicalism is based ultimately on a kind of idealism: the idealism of objectivity. Objectivity is not reality. It is just one way of understanding reality. Still, even if objective understanding can be only partial, it is worth trying to extend it, for a simple reason.

The pursuit of an objective understanding of reality is still the only way to expand our knowledge of what there is beyond the way it appears to us. Even if we have to acknowledge the reality of some things that we can't grasp objectively, as well as the ineliminable subjectivity of some aspects of our own experience which we can grasp *only* subjectively, the pursuit of an objective concept of mind is simply part of the general pursuit of understanding. To give it up because it cannot be complete would be like giving up axiomatization in mathematics because it cannot be complete.

10. I want now to change the subject slightly, from the general concept of mind to the individual concept of the self.[3]

Even if we accept the liberal realist picture of the world that

---

[3] I discuss this problem more fully in "The Objective Self," forthcoming in *Knowledge and Mind* (essays in honor of Norman Malcolm), Carl Ginet and Sydney Shoemaker, eds.

I have sketched, as a place that contains us, with all our perspectives, and other beings, with all theirs, something still remains puzzling. Each of us will have to admit that one of the most important things about this centerless world is that one of the persons in it is himself. What kind of a fact is that? What kind of fact is it, for example, that I am Thomas Nagel? It is a fact I appear to recognize whenever I take a step back from my current viewpoint to pursue a more objective overview, which includes that person. But the fact that I am TN does not fit very easily into the conception of the world from nowhere within it that this detachment is supposed to produce. How can there be room in such an impersonally described world for the further fact that TN is me, the locus of *my* consciousness?

I am not a solipsist. I do not believe that the point of view from which I see the world is *the* perspective of reality. The world is seen from many points of view, including this one; there are many subjects of consciousness in it, and an adequate center less conception of the world must include them all. If the world really doesn't have a particular point of view, there can be no irreducibly first-person facts. On the other hand nothing could be clearer than that I *am* TN, and this seems like a first-person fact if there ever was one. So I seem to have on my hands a fact about the world which both must exist (for how things are would be incomplete without it) and cannot exist (for how things are cannot include it).

A full treatment of this question would have to deal with the charge that it is a pseudo-problem caused by misunderstanding of the logic of token-reflexives. I believe this charge can be shown to be false. But instead of trying to do that here, I am going to propose a solution to the problem. Any problem that has a solution must be real.

The philosophical thought that I am TN has a content very different from what is conveyed when I use the words 'I am TN' to *introduce* myself to you. But it is a content for which room can

be found in a centerless conception of the world. A similar fact can be discovered about each of you, which is the object of your thought that you are the particular human individual you in fact are. But I shall proceed in the Cartesian first person, which is intended to be understood by each of you as applying to himself.

When I think that I am TN, I think that the real me regards the world from the point of view of TN: *occupies* TN, so to speak. Another way of putting it would be to say that the publicly identifiable person TN contains the real me. But what is this 'real me'? Why isn't the real me just TN*?*

Strange as it may sound, I find it extremely puzzling that I should be TN, and I think if we can understand what is puzzling about it, we will get to the heart of the problem. Actually the question "How can I be TN?" has two aspects, corresponding to the two directions in which it can be asked. So far I have been discussing the question "What kind of fact is it that the particular person TN is *me*?" But there is also the question "How can *I* be the particular person TN, or any other particular person for that matter? How can *I* be *somebody*?" It is this question that is really basic.

Think about the world as a whole, with the publicly identifiable person you are as one of its contents, and ask yourself, "How can I be anything so *specific* as that; how can I be *merely* a particular person?" This is different from the original question, which was, "What makes that person *me*?" The problem here is how I can be *anything* so specific as a particular person in the world *at all* — *any* person.

The trouble is that my connection with TN seems *arbitrary.* When I consider the world as a whole, as existing from no particular point of view, TN is just one person among many others, and although on Earth his species is dominant, that fades to insignificance on an astronomical scale. How can I, who am thinking about the entire, centerless universe, be anything so specific as *this:* this creature *in* the universe, existing in a tiny morsel of

space and time, with a definite and by no means universal mental and physical organization?

I believe the solution to both these problems is the same. The problems are: (1) What kind of fact is it that TN is *me*? (2) How can *I* be anything so specific as TN? And the solution comes from identifying the 'I' in these thoughts with an objective self that each of us contains, that has an unlimited capacity to step back from the standpoint of the person I am in order to form a new conception of the world in which that person and his states are located.

The picture is this. Essentially I am a subject that apprehends a centerless world. Essentially I have no particular point of view. In fact I ordinarily view the world through the eyes, the person, the daily life of TN, as through a window. But the experiences and the perspective of TN with which I am directly presented are not essential to the point of view of the true self. The true self apprehends the world from no point of view and includes in its conception TN and his perspective among the *contents* of the world.

How do I separate my true self from this *person*? By treating the experiences of this person, which depend on his particularity, as data. I throw him into the world as a thing that interacts with the rest of it, and ask what the world must be like from *no* point of view in order to appear to him as it does from *his* point of view. I can reason in this way about anyone else as well as about him. Even though I receive the information of his point of view directly, I try to deal with it for the purpose of constructing a realist and partly objective picture in a way similar to that which would be appropriate if the information were coming to me indirectly.

So when I have the philosophical thought that I am TN, I am recognizing that the particular objective self that is the subject of this centerless conception of a world in which TN is located, is also viewing the world from within through the perspective of

TN. And when I am philosophically puzzled over how I can be *merely* the person TN, it is because I am thinking of myself as the objective self which occupies TN. The paradox disappears because TN, like the rest of you, turns out to be not merely a particular creature with a very specific perspective on the world from his position inside it. Any human being also contains a very different kind of subject, largely undeveloped or unexplored, but with the potential for indefinite further impersonal and objective apprehension of the world. The 'further fact' that I am TN is the fact that this impersonal conception of the world can close over itself by locating the subject that forms it at a particular point in the world that it apprehends. It is attached to, and developed from, the perspective of TN. And since that is not an irreducibly first-person fact, it can be part of the real world.

11. Let me close by saying something about the more general implications of these remarks, which have been concerned specifically with the philosophy of mind. In trying to explain how the mind and the self are to be included in the real world that simply exists, I have distinguished between reality and objective reality, and also between objectivity and particular conceptions of objectivity. The physical conception of objectivity is inappropriate for increasing our understanding of the mind; and even the kind of objectivity that is appropriate for this purpose will not permit us to form a complete idea of all the various incompatible mental perspectives. The general upshot, that applies to ethics as well, is that one should pursue the kind of objectivity appropriate to the subject one is trying to understand, and that even the right kind of objectivity may not exhaust the subject completely.

The problem of bringing together subjective and objective views of the world can be approached from either direction. If one starts from the subjective side the problem is the traditional one of skepticism, idealism, or solipsism. How, given my personal experiential perspective, can I form a conception of the world as it is

independent of my perception of it? And how can I know that this conception is correct? (The question may also be asked from the point of view of the collective human perspective rather than from that of an individual.) If on the other hand one starts from the objective side, the problem is how to accommodate, in a world that simply exists from no perspective, any of the following things: (a) oneself, (b) one's point of view, (c) the point of view of other selves, similar and dissimilar, and (d) the objects of various types of judgment that seem to emanate from these perspectives.

It is this second version of the problem that interests me. It is the obverse of skepticism because the *given* is objective reality — or the idea of an objective reality — and what is problematic by contrast is subjective reality. Without receiving full acknowledgment this approach has been very influential in recent analytic philosophy. It accords well with a bias toward physical science as a paradigm of understanding.

But if under the pressure of realism we admit that there are things which cannot be understood in this way, then other ways of understanding them must be sought. One way is to enrich the notion of objectivity. But to insist in every case that the most objective and detached account of a phenomenon is the correct one is likely to lead to reductive conclusions. Sometimes, in the philosophy of mind, and as I hope to show, in ethics, the truth is not to be found by traveling as far away from one's personal perspective as possible.

II. VALUE

1. Whether values can be objective depends on whether an interpretation of objectivity can be found that allows us to advance our knowledge of what to do, what to want, and what things provide reasons for and against action. Last week I argued that the physical conception of objectivity was not able to provide an understanding of the mind, but that another conception was available which allowed external understanding of at least some *aspects* of mental phenomena. A still different conception is required to make sense of the objectivity of values, for values are neither physical nor mental. And even if we find a conception, it must be applied with care. Not all values are likely to prove to be objective in any sense.

Let me say in advance that my discussion of values and reasons in this lecture and the next will be quite general. I shall be talking largely about what determines whether something has value, or whether someone has a reason to do or want something. I shall say nothing about how we pass from the identification of values and reasons to a conclusion as to what should be *done*. That is of course what makes reasons important; but I shall just assume that values do often provide the basis for such conclusions, without trying to describe even in outline how the full process of practical reasoning works. I am concerned here only with the general question, whether values have an objective foundation at all.

In general, as I said last time, objectivity is advanced when we step back, detach from our earlier point of view toward something, and arrive at a new view of the whole that is formed by including ourselves and our earlier viewpoint in what is to be understood.

In theoretical reasoning this is done by forming a new conception of reality that includes ourselves as components. This involves an alteration, or at least an extension, of our beliefs. Whether the effort to detach will actually result in an increase of understanding depends on the creative capacity to form objective ideas which is called into action when we add ourselves to the world and start over.

In the sphere of values or practical reasoning, the problem is somewhat different. As in the theoretical case, in order to pursue objectivity we must take up a new, comprehensive viewpoint after stepping back and including our former perspective in what is to be understood. But in this case the new viewpoint will be *not* a new set of *beliefs,* but a new, or extended, set of *values.* If objectivity means anything here, it will mean that when we detach from our individual perspective and the values and reasons that seem acceptable from within it, we can sometimes arrive at a new conception which may endorse some of the original reasons but will reject some as subjective appearances and add others. This is what is usually meant by an objective, disinterested view of a practical question.

The basic step of placing ourselves and our attitudes within the world to be considered is familiar, but the form of the result — a new set of values, reasons, and motives — is different. In order to discover whether there are any objective values or reasons we must try to arrive at *normative* judgments, with *motivational content,* from an impersonal standpoint: a standpoint outside of our lives. We cannot use a *non-normative* criterion of objectivity: for *if* any values are objective, they are objective *values,* not objective anything else.

2. There are many opinions about whether what we have reason to do or want can be determined from a detached standpoint toward ourselves and the world. They range all the way from the view that objectivity has *no* place in this domain except what is

inherited from the objectivity of those theoretical and factual elements that play a role in practical reasoning, to the view that objectivity applies here, but with a nihilistic result: i.e., that nothing is objectively right or wrong because objectively nothing matters. In between are many positive objectifying views which claim to get some definite results from a detached standpoint. Each of them is criticized by adherents of opposing views either for trying to force too much into a single objective framework or for according too much or too little respect to divergent subjective points of view.

Here as elsewhere there is a direct connection between the goal of objectivity and the belief in *realism.* The most basic idea of practical objectivity is arrived at by a practical analogue of the rejection of solipsism or idealism in the theoretical domain. Just as realism about the facts leads us to seek a detached point of view from which reality can be discerned and appearance corrected, so realism about values leads us to seek a detached point of view from which it will be possible to correct inclination and to discern what we really should do, or want. Practical objectivity means that practical reason can be understood and even engaged in by the objective self.

This assumption, though powerful, is not yet an ethical position. It merely marks the place which an ethical position will occupy if we can make any sense of the subject. It says that the world of reasons, including my reasons, does not exist only from my point of view. I am in a world whose properties are to a certain extent independent of what I think, and if I have reasons to act it is because the person who I am has those reasons, in virtue of his condition and circumstances. One would expect those reasons to be understandable from outside. Here as elsewhere objectivity is a form of understanding not necessarily available for all of reality. But it is reasonable at least to look for such understanding over as wide an area as possible.

3. It is important not to lose sight of the dangers of *false* objectification, which too easily elevate personal tastes and prejudices into cosmic values. But initially, at least, it is natural to look for some objective account of those reasons that appear from one's own point of view.

In fact those reasons usually present themselves with some pretensions of objectivity to begin with, just as perceptual appearances do. When two things look the same size to me, they look at least initially as if they *are* the same size. And when I want to take aspirin because it will cure my headache, I believe at least initially that this *is* a reason for me to take aspirin, that it can be recognized as a reason from outside, and that if I failed to take it into account, that would be a mistake, and others could recognize this.

The ordinary process of deliberation, aimed at finding out what I have reason to do, assumes that the question has an answer. And in difficult cases especially, deliberation is often accompanied by the belief that I may not *arrive* at that answer. I do not assume that the correct answer is just whatever will result or has resulted from consistent application of deliberative methods — even assuming perfect information about the facts. In deliberation we are trying to arrive at conclusions that are correct in virtue of something *independent* of our arriving at them. If we arrive at a conclusion, we believe that it would have been correct even if we *hadn't* arrived at it. And we can also acknowledge that we might be *wrong,* since the process of reasoning doesn't guarantee the correctness of the result. So the pursuit of an objective account of practical reasons has its basis in the realist claims of ordinary practical reasoning. In accordance with pretheoretical judgment we adopt the working hypothesis that there are reasons which may diverge from actual motivation even under conditions of perfect information — as reality can diverge from appearance — and then consider what form these reasons take. I shall say more about the general issue of realism later on. But first I want to

concentrate on the process of thought by which, against a realist background, one might try to arrive at objective conclusions about reasons for action. In other words, if there really are values, how is objective knowledge of them possible?

In this inquiry no particular hypothesis occupies a privileged position, and it is certain that some of our starting points will be abandoned as we proceed. However, one condition on reasons obviously presents itself for consideration: a condition of generality. This is the condition that if something provides a reason for a particular individual to do something, then there is a general form of that reason which applies to anyone else in comparable circumstances. What count as comparable circumstances depends on the general form of the reason. This condition is not tautological. It is a rather strong condition which may be false, or true only for some kinds of reasons. But the search for generality is a natural beginning.

4. There is more than one type of generality, and no reason to assume that a single form will apply to every kind of reason or value. In fact I think that the choice among types of generality defines some of the central issues of contemporary moral theory.

One respect in which reasons may vary is in their *breadth.* A general principle may apply to everyone but be quite specific in content, and it is an open question to what extent narrower principles of practical reasons (don't lie; develop your talents) can be subsumed under broader ones (don't hurt others; consider your long-term interests), or even at the limit under a single widest principle from which all the rest derive. Reasons may be general, in other words, without forming a unified system that always provides a method for arriving at determinate conclusions about what one should do.

A second respect in which reasons vary is in their *relativity to the agent,* the person for whom they are reasons. The distinction between reasons that are relative to the agent and reasons that

are not is an extremely important one. I shall follow Derek Parfit in using the terms 'agent-relative' and 'agent-neutral' to mark this distinction. (Formerly I used the terms 'subjective' and 'objective,' but those terms are here reserved for other purposes.)

If a reason can be given a general form which does *not* include an essential reference to the person to whom it applies, it is an *agent-neutral* reason. For example, if it is a reason for *anyone* to do or want something that it would reduce the amount of wretchedness in the world, then that is an agent-neutral reason.

If on the other hand the general form of a reason *does* include an essential reference to the person to whom it applies, it is an *agent-relative* reason. For example, if it is a reason for anyone to do or want something that it would be in *his* interest, then that is an agent-relative reason. In such a case, if something were in Jones's interest but contrary to Smith's, Jones would have reason to want it to happen and Smith would have the *same* reason to want it *not* to happen. (Both agent-relative and agent-neutral reasons are objective, since both can be understood from outside the viewpoint of the individual who has them.)

A third way in which reasons may vary is in their degree of externality, or independence of the interests of sentient beings. Most of the apparent reasons that initially present themselves to us are intimately connected with interests and desires, our own or those of others, and often with experiential satisfaction. But it is conceivable that some of these interests give evidence that their objects have intrinsic value independent of the satisfaction that anyone may derive from them or of the fact that anyone wants them — independent even of the existence of beings who can take an interest in them. I shall call a reason *internal* if it depends on the existence of an interest or desire in someone, and *external* if it does not. External reasons were believed to exist by Plato, and more recently by G. E. Moore, who believed that aesthetic value provided candidates for this kind of externality.

These three types of variation cut across one another. For-

mally, a reason may be narrow, external, and agent-relative (don't eat pork, keep your promises), or broad, internal, and agent-neutral (promote happiness), or internal and agent-relative (promote your own happiness). There may be other significant dimensions of variation. I want to concentrate on these because they locate the main controversies about what ethics is. Reasons and values that can be described in these terms provide the material for objective judgments. If one looks at human action and its conditions from outside and considers whether some normative principles are plausible, these are the forms they will take.

The actual *acceptance* of a general normative judgment will have motivational implications, for it will commit you under some circumstances to the acceptance of reasons to want and do things *yourself.*

This is most clear when the objective judgment is that something has *agent-neutral* value. That means *anyone* has reason to want it to happen — *and that includes someone considering the world in detachment from the perspective of any particular person within it.* Such a judgment has motivational content even before it is brought back down to the particular perspective of the individual who has accepted it objectively.

Agent-relative reasons are different. An objective judgment that some kind of thing has *agent-relative* value commits us only to believing that someone has reason to want and pursue it if it is related to him in the right way (being in *his* interest, for example). Someone who accepts this judgment is not committed to wanting it to *be the case* that people in general are influenced by such reasons. The judgment commits him to wanting something only when its implications are drawn *for the individual person he happens to be.* With regard to others, the content of the objective judgment concerns only what *they* should do or want.

I believe that judgments of both these kinds, as well as others, are evoked from us when we take up an objective standpoint. And I believe such judgments can be just as true and compelling

as objective factual judgments about the real world that con-
tains us.

5. When we take the step to objectivity in practical reasoning
by detaching from our own point of view, the question we must
ask ourselves is this: What reasons for action can be said to apply
to people when we regard them from a standpoint detached from
the values of any particular person?

The simplest answer, and one that some people would give, is
"None." But that is not the only option. The suggested classi-
fication of types of generality provides a range of alternative
hypotheses. It also provides some flexibility of response, for with
regard to any reason that may appear to a particular individual
to exist subjectively, the corresponding objective judgment may be
that it does not exist at all, or that it corresponds to an agent-
neutral, external value, or anything in between.

The choice among these hypotheses, plus others not yet imag-
ined, is difficult, and there is no general method of making it any
more than there is a general method of selecting the most plau-
sible objective account of the facts on the basis of the appearances.
The only 'method,' here or elsewhere, is to try to generate hypoth-
eses and then to consider which of them seems most reasonable, in
light of everything else one is fairly confident of.

This is not quite empty, for it means at least that logic alone
can settle nothing. We do *not* have to be shown that the denial
of some kind of objective values is *self-contradictory* in order to
be reasonably led to accept their existence. There is no constraint
to pick the weakest or narrowest or most economical principle
consistent with the initial data that arise from individual perspec-
tives. Our admission of reasons beyond these is determined *not*
by logical entailment, but by what we cannot help believing, or at
least finding most plausible among the alternatives.

In this respect it is no different from anything else: theoretical
knowledge does not arise by deductive inference from the appear-

ances either. The main difference is that our objective thinking about practical reasons is very primitive, and has difficulty taking even the first step. Philosophical skepticism and idealism about values are much more popular than their metaphysical counterparts. Nevertheless I believe they are no more correct. I shall argue that although no *single* objective principle of practical reason like egoism or utilitarianism covers everything, the acceptance of some objective values is unavoidable —not because the alternative is inconsistent but because it is not *credible.* Someone who, as in Hume's example, prefers the destruction of the whole world to the scratching of his finger, may not be involved in a *contradiction* or in any false *expectations,* but he is unreasonable nonetheless (to put it mildly), and anyone else not in the grip of an overly narrow conception of what reasoning is would regard his preference as objectively wrong.

6. But even if it is unreasonable to deny that anyone ever objectively has a reason to do anything, it is not easy to find positive objective principles that *are* reasonable. I am going to attempt to defend a few in the rest of this lecture and the next. But I want to acknowledge in advance that it is not easy to follow the objectifying impulse without distorting individual life and personal relations. We want to be able to understand and accept the way we live from outside, but it may not always follow that we should control our lives from inside by the terms of that external understanding. Often the objective viewpoint will not be suitable as a replacement for the subjective, but will coexist with it, setting a standard with which the subjective is constrained not to clash. In deciding what to do, for example, we should not reach a result different from what we could decide objectively that that *person* should do —but we need not arrive at the result in the same way from the two standpoints.

Sometimes, also, the objective standpoint will allow us to judge how people should *be* or should *live,* without permitting us

to translate this into a judgment about what they have *reasons* to do. For in some respects it is better to live and act not for reasons, but because we cannot help it. This is especially true of close personal relations. Here the objective standpoint cannot be brought into the perspective of *action* without destroying precisely what it affirms the value of. Nevertheless the possibility of this objective affirmation is important. We should be *able* to view our lives from outside without extreme dissociation or distaste, and the extent to which we should live without *considering* the objective point of view or even any reasons *at all* is itself *determined* largely from that point of view.

It is also possible that some idiosyncratic individual grounds of action, or the values of strange communities, will prove objectively inaccessible. To take an example in our midst: I don't think that people who want to be able to run twenty-six miles without stopping are irrational, but their reasons can be understood only from the perspective of a value system that is completely alien to me, and will I hope remain so. A correct objective view will have to allow for such pockets of unassimilable subjectivity, which need not clash with objective principles but won't be affirmed by them either. Many aspects of personal taste will come in this category, if, as I think, they cannot all be brought under a general hedonistic principle.

But the most difficult and interesting problems of accommodation appear where objectivity *can* be employed as a standard, but we have to decide *how*. Some of the problems are these: To what extent should an objective view admit *external* values? To what extent should it admit *internal* but *agent-neutral* values? To what extent should the reasons to respect the interests of *others* take an *agent-relative* form? To what extent is it legitimate for each person to give priority to his own interests? These are all questions about the proper form of generality for different kinds of practical reasoning, and the proper relation between objective principles and the deliberations of individual agents. I shall return to some of

them later, but there is a great deal that I shall not get to.

I shall not, for example, discuss the question of external values, i.e., values which may be *revealed* to us by the attractiveness of certain things, but whose existence is independent of the existence of any interests or desires. I am not sure whether there are any such values, though the objectifying tendency produces a strong impulse to believe that there are, especially in aesthetics where the object of interest is external and the interest seems perpetually capable of criticism in light of further attention to the object.

What I shall discuss is the proper form of *internal* values or reasons —those which depend on interests or desires. They can be objectified in more than one way, and I believe different forms of objectification are appropriate for different cases.

7. I plan to take up some of these complications in the next lecture. Let me begin, however, with a case for which I think the solution is simple: that of pleasure and pain. I am not an ethical hedonist, but I think pleasure and pain are very important, and they have a kind of neutrality that makes them fit easily into ethical thinking —unlike preferences or desires, for example, which I shall discuss later on.

I mean the kinds of pleasure and pain that do not depend on activities or desires which *themselves* raise questions of justification and value. Many pleasures and pains are just sensory experiences in relation to which we are fairly passive, but toward which we feel involuntary desire or aversion. Almost everyone takes the avoidance of his own pain and the promotion of his own pleasure as subjective reasons for action in a fairly simple way; they are not backed up by any further reasons. On the other hand if someone pursues pain or avoids pleasure, these idiosyncracies usually *are* backed up by further reasons, like guilt or sexual masochism. The question is, what sort of general value, if any, ought to be assigned to pleasure and pain when we consider these facts from

an objective standpoint?

It seems to me that the least plausible hypothesis is the zero position, that pleasure and pain have no value of any kind that can be objectively recognized. That would mean that looking at it from outside, you couldn't even say that someone had a reason not to put his hand on a hot stove. Try looking at it from the outside and see whether you can manage to withhold that judgment.

But I want to leave this position aside, because what really interests me is the choice between two other hypotheses, both of which admit that people have reason to avoid their own pain and pursue their own pleasure. They are the fairly obvious general hypotheses formed by assigning (a) agent-relative or (b) agent-neutral value to those experiences. If the avoidance of pain has only agent-relative value, then people have reason to avoid their own pain, but not to avoid the pain of others (unless other kinds of reasons come into play). If the avoidance of pain has agent-neutral value as well, then *anyone* has a reason to want *any* pain to stop, whether or not it is his. From an objective standpoint, which of these hypotheses is more plausible? Is the value of sensory pleasure and pain agent-relative or agent-neutral?

I believe it is agent-neutral, at least in part. That is, I believe pleasure is a good thing and pain is a bad thing, and that the most reasonable objective principle which admits that each of us has reason to pursue his own pleasure and avoid his own pain will acknowledge that these are not the only reasons present. This is a normative claim. Unreasonable, as I have said, does not mean inconsistent.

In arguing for this claim, I am somewhat handicapped by the fact that I find it self-evident. It is therefore difficult for me to find something still more certain with which to back it up. But I shall try to say what is wrong with rejecting it, and with the reasons that may lie behind its rejection. What would it be to really *accept* the alternative hypothesis that pleasure and pain are not impersonally good or bad? If I accept this hypothesis, assuming

at the same time that each person has reason to seek pleasure and avoid pain for *himself,* then when I regard the matter objectively the result is very peculiar. I will have to believe that I have a reason to take aspirin for a headache, but that there *is no reason* for me to *have* an aspirin. And I will have to believe the same about anyone else. From an objective standpoint I must judge that everyone has reason to pursue a type of result that is *impersonally valueless,* that has value only to *him*.

This needs to be explained. If agent-neutral reasons are not ruled out of consideration from the start (and one would need reasons for that), why do we not have evidence of them here? The avoidance of pain is not an individual project, expressing the agent's personal values. The desire to make pain stop is simply *evoked* in the person who feels it. He may decide for various reasons not to stop it, but in the first instance he doesn't have to *decide* to want it to stop: he just does. He wants it to go away because it's *bad:* it is not *made* bad by his deciding that he wants it to go away. And I believe that when we think about it objectively, concentrating on what pain is *like,* and ask ourselves whether it is (a) not bad at all, (b) bad only for its possessor, or (c) bad *period,* the third answer is the one that needs to be argued *against,* not the one that needs to be argued *for*. The philosophical problem here is to get rid of the obstacles to the admission of the obvious. But first they have to be identified.

Consider how *strange* is the question posed by someone who wants a justification for altruism about such a basic matter as this. Suppose he and some other people have been admitted to a hospital with severe burns after being rescued from a fire. "I understand how *my* pain provides *me* with a reason to take an analgesic," he says, "and I understand how my groaning neighbor's pain gives *him* a reason to take an analgesic; but how does *his* pain give *me* any reason to want him to be given an analgesic? How can *his* pain give *me* or anyone else looking at it from outside a reason?"

This question is *crazy.* As an expression of puzzlement, it has that characteristic philosophical craziness which indicates that something very fundamental has gone wrong. This shows up in the fact that the *answer* to the question is *obvious,* so obvious that to ask the question is obviously a philosophical act. The answer is that pain is *awful.* The pain of the man groaning in the next bed is just as awful as yours. That's your reason to want him to have an analgesic.

Yet to many philosophers, when they think about the matter theoretically, this answer seems not to be available. The pain of the person in the next bed is thought to need major external help before it can provide me with a reason for wanting or doing anything: otherwise it can't get its hooks into me. Since most of these people are perfectly aware of the force such considerations actually have for them, justifications of some kind are usually found. But they take the form of working *outward* from the desires and interests of the individual for whom reasons are being sought. The burden of proof is thought always to be on the claim that he has reason to care about anything that is not *already* an object of his interest.

These justifications are unnecessary. They plainly falsify the real nature of the case. My reason for wanting my neighbor's pain to cease is just that it's awful, and I know it.

8. What is responsible for this demand for justification with its special flavor of philosophical madness? I believe it is something rather deep, which doesn't surface in the ordinary course of life: an inappropriate sense of the burden of proof. Basically, we are being asked for a demonstration of the *possibility* of real impersonal values, on the assumption that they are *not* possible unless such a general proof can be given.

But I think this is wrong. We can already *conceive* of such a possibility, and once we take the step of thinking about what reality, if any, there is in the domain of practical reason, it be-

comes a possibility we are bound to consider, that we cannot help considering. If there really are *reasons* not just motivational pushes and pulls, and if agent-neutral reasons are among the kinds we can conceive of, then it becomes an obvious possibility that physical pain is simply bad: that even from an impersonal stand-point there is reason to want it to stop. When we view the matter objectively, this is one of the general positions that naturally sug-gests itself.

And once this is seen as a *possibility,* it becomes difficult *not* to accept it. It becomes a hypothesis that has to be *dislodged* by anyone who wishes to claim, for example, that all reasons are agent-relative. The question is, what are the alternatives, once we take up the objective standpoint? We must think *something.* If there is room in the realistic conception of reasons for agent-neutral values, then it is unnatural *not* to ascribe agent-neutral badness to burn pains. That is the natural conclusion from the fact that any-one who has a burn pain and is therefore closest to it wants acutely to be rid of it, and requires no indoctrination or training to want this. This evidence does not *entail* that burn pains are im-personally bad. It is logically conceivable that there is nothing bad about them at all, or that they provide only agent-relative rea-sons to their possessors to want them to go away. But to take such hypotheses seriously we would need justifications of a kind that seem totally unavailable in this case.

What could possibly show us that acute physical pain, which everyone finds horrible, is in reality not impersonally bad at all, so that except from the point of view of the sufferer it doesn't in itself matter? Only a very remarkable and farfetched picture of the value of a cosmic order beyond our immediate grasp, in which pain played an essential part which made it good or at least neutral —or else a demonstration that there can *be* no agent-neutral values. But I take it that neither of these is available: the first because the Problem of Evil has not been solved, the second because the absence of a logical demonstration that there *are*

agent-neutral values is not a demonstration that there are *not* agent-neutral values.

My position is this. No demonstration is necessary in order to allow us to *consider* the possibility of agent-neutral reasons: the possibility simply *occurs* to us once we take up an objective stance. And there is no mystery about how an individual could have a reason to want something independently of its relation to his particular interests or point of view, because beings like ourselves are not *limited* to the particular point of view that goes with their personal position inside the world. They are also, as I have put it earlier, *objective selves:* they cannot *help* forming an objective conception of the world with themselves in it; they cannot help trying to arrive at judgments of *value* from that standpoint; they cannot help asking whether, from that standpoint, in abstraction from who in the world they are, they have any reason to want anything to be the case or not — any reason to want anything to happen or not.

Agent-neutral reasons do not have to find a miraculous source in our personal lives, because we are not *merely* personal beings: we are also importantly and essentially viewers of the world *from nowhere within it* — and in this capacity we remain open to judgments of value, both general and particular. The possibility of agent-neutral values is evident as soon as we begin to think from this standpoint about the reality of any reasons whatever. If we acknowledge the possibility of realism, then we cannot rule out agent-neutral values in advance.

Realism is therefore the fundamental issue. If there really are values and reasons, then it should be possible to expand our understanding of them by objective investigation, and there is no reason to rule out the natural and compelling objective judgment that pain is impersonally bad and pleasure impersonally good. So let me turn now to the abstract issue of realism about values.

9. Like the presumption that things exist in an external world,

the presumption that there are real values and reasons can be defeated in individual cases, if a purely subjective account of the appearances is more plausible. And like the presumption of an external world, its complete falsity is not self-contradictory. The reality of values, agent-neutral or otherwise, is not *entailed* by the totality of appearances any more than the reality of a physical universe is. But if either of them is recognized as a possibility, then its reality in detail can be confirmed by appearances, at least to the extent of being rendered more plausible than the alternatives. So a lot depends on whether the possibility of realism is admitted in the first place.

It is very difficult to argue for such a possibility. Sometimes there will be arguments against it, which one can try to refute. Berkeley's argument against the conceivability of a world independent of experience is an example. But what is the result when such an argument is refuted? Is the possibility in a stronger position? I believe so: in general, there is no way to prove the possibility of realism; one can only refute impossibility arguments, and the more often one does this the more confidence one may have in the realist alternative. So to consider the merits of an admission of realism about value, we have to consider the reasons against it. I shall discuss three. They have been picked for their apparent capacity to convince people.

The first argument depends on the question-begging assumption that if values are real, they must be real objects of some other kind. John Mackie, for example, in his recent book *Ethics,* denies the objectivity of values by saying that they are not part of the fabric of the world, and that if they were, they would have to be "entities or qualities or relations of a very strange sort, utterly different from anything else in the universe." [4] Apparently he has a very definite picture of what the universe is like, and assumes

---

[4] J. L. Mackie, *Ethics* (Harmondsworth, Middlesex: Penguin Books, 1977), p. 38.

that realism about value would require crowding it with extra
entities, qualities, or relations —things like Platonic Forms or
Moore's non-natural qualities. But this assumption is not correct.
The impersonal badness of pain is not some mysterious further
property that all pains have, but just the fact that there is reason
for anyone capable of viewing the world objectively to want it to
stop, whether it is his or someone else's. The view that values are
real is not the view that they are real occult entities or properties,
but that they are real *values:* that our claims about value and
about what people have reason to do may be *true* or *false* inde-
pendently of our beliefs and inclinations. No *other* kinds of truths
are involved. Indeed, no other kinds of truths *could* imply the
reality of values.[5]

The second argument I want to consider is not, like the first,
based on a misinterpretation of moral objectivity. Instead, it tries
to represent the unreality of values as an objective *discovery.* The
argument is that if claims of value have to be objectively correct
or incorrect, and if they are not reducible to any *other* kind of
objective claim, then we can just *see* that all positive value claims
must be false. Nothing has any objective value, because objec-
tively nothing matters at all. If we push the claims of objective
detachment to their logical conclusion, and survey the world from
a standpoint completely detached from all interests, we discover
that there is *nothing* —no values left of any kind: things can be

---

5 In discussion, Mackie claimed that I had misrepresented him, and that his
disbelief in the reality of values and reasons does not depend on the assumption
that to be real they must be strange *entities* or *properties*. As he says in his book,
it applies directly to reasons themselves. For whatever they are they are not needed
to explain anything that happens, and there is consequently no reason to believe in
their existence. But I would reply that this raises the same issue. It begs the ques-
tion to assume that *explanatory* necessity is the test of reality in this area. The
claim that certain reasons exist is a normative claim, not a claim about the best
explanation of anything. To assume that only what has to be included in the best
explanatory picture of the world is real, is to assume that there are no irreducibly
normative truths.

There is much more to be said on both sides of this issue, and I hope I have
not misrepresented Mackie in this short footnote.

said to matter at all only to individuals within the world. The result is objective nihilism.

I don't deny that the objective standpoint tempts one in this direction. But I believe this can seem like the required conclusion only if one makes the mistake of assuming that objective judgments of value must emerge from the detached standpoint *alone*. It is true that with nothing to go on but a conception of the world from nowhere, one would have no way of telling whether anything had value. But an objective view has more to go on, for its data include the appearance of value to individuals with particular perspectives, including oneself. In this respect practical reason is no different from anything else. Starting from a pure idea of a possible reality and a very impure set of appearances, we try to fill in the idea of reality so as to make some partial sense of the appearances, using objectivity as a method. To find out what the world is like from outside we have to approach it from within: it is no wonder that the same is true for ethics. And indeed, when we take up the objective standpoint, the problem is not that values seem to disappear but that there seem to be too many of them, coming from every life and drowning out those that arise from our own. It is just as easy to form desires from an objective standpoint as it is to form beliefs. Probably easier. Like beliefs, these desires and evaluations must be criticized and justified partly in *terms* of the appearances. But they are not just further appearances, any more than the beliefs about the world which arise from an impersonal standpoint are just further appearances.

The third type of argument against the objective reality of values is an empirical argument. It is also perhaps the most common. It is intended not to rule out the possibility of real values from the start, but rather to demonstrate that even if their *possibility* is admitted, we have no reason to believe that there are any. The claim is that if we consider the wide cultural variation in normative beliefs, the importance of social pressure and other psychological influences to their formation, and the difficulty

of settling moral disagreements, it becomes highly implausible that they are anything but pure appearances.

Anyone offering this argument must admit that not every psychological factor in the explanation of an appearance shows that the appearance corresponds to nothing real. Visual capacities and elaborate training play a part in explaining the physicist's perception of a cloud-chamber track, or a student's coming to believe a proposition of geometry, but the path of the particle and the truth of the proposition also play an essential part in these explanations. So far as I know, no one has produced a general account of the kinds of psychological explanation that discredit an appearance. But some skeptics about ethics feel that because of the way we acquire moral beliefs and other impressions of value, there are grounds for confidence that no real, objective values play a part in the explanation.

I find the popularity of this argument surprising. The fact that morality is socially inculcated and that there is radical disagreement about it across cultures, over time, and even within cultures at a time is a poor reason to conclude that values have no objective reality. Even where there is truth, it is not always easy to discover. Other areas of knowledge are taught by social pressure, many truths as well as falsehoods are believed without rational grounds, and there is wide disagreement about scientific and social facts, especially where strong interests are involved which will be affected by different answers to a disputed question. This last factor is present throughout ethics to a uniquely high degree: it is an area in which one would expect extreme variation of belief and radical disagreement however objectively real the subject actually was. For comparably motivated disagreements about matters of fact, one has to go to the heliocentric theory, the theory of evolution, the Dreyfus case, the Hiss case, and the genetic contribution to racial differences in I.Q.

Although the methods of ethical reasoning are rather primitive, the degree to which agreement can be achieved and social

prejudices transcended in the face of strong pressures suggests that something real is being investigated, and that part of the explanation of the appearances, both at simple and at complex levels, is that we perceive, often inaccurately, that certain reasons for action exist, and go on to infer, often erroneously, the general form of the principles that best accounts for those reasons.

The controlling conception that supports these efforts at understanding, in ethics as in science, is realism, or the possibility of realism. Without being sure that we will find one, we look for an account of what reasons there really are, an account that can be objectively understood.

I have not discussed all the possible arguments against realism about values, but I have tried to give general reasons for skepticism about such arguments. It seems to me that they tend to be supported by a narrow preconception of what there *is,* and that this is essentially question-begging.

10. Let me close this lecture by indicating what I plan to discuss next week. So far I have been arguing against skepticism, and in favor of realism and the pursuit of objectivity in the domain of practical reason. But if realism is admitted as a possibility, one is quickly faced with the opposite of the problem of skepticism. This is the problem of over-objectification: the temptation to interpret the objectivity of reasons in too strong and unitary a way.

In ethics, as in metaphysics, the allure of objectivity is very great: there is a persistent tendency in both areas to seek a *single, complete* objective account of reality — in the area of value that means a search for the most objective possible account of all reasons for action: the account acceptable from a maximally detached standpoint.

This idea underlies the fairly common moral assumption that the only real values are agent-neutral values, and that someone can really have a reason to do something only if there is an agent-neutral

reason for it to *happen.* That is the essence of consequentialism: the only reason for anyone to do anything is that it would be better in itself, considering the world as a whole, if he did it. (The idea also finds a reflection in Professor Hare's view about the only kind of judgment that moral language can be used to express: for his claim that moral judgments are universally *prescriptive* means that they depend on what one would *want to happen,* considering the question from all points of view —rather than on what one would think people had *reason to do,* considering the question in this way. Consequently, any principle that was *moral* in his sense would have to be *agent-neutral.)*

In the next lecture I shall try to explain why ethics has to be based not only on agent-neutral values like those that attach to pleasure and pain. We can no more assume that all reasons are agent-neutral than that all reality is physical. I argued earlier that not everything there is can be gathered into a uniform conception of the universe from nowhere within it. If certain perspectives evidently exist which cannot be analyzed in physical terms, we must modify our idea of objective reality to include them. If that is not enough, we must admit to reality some things that cannot be objectively understood. Similarly, if certain reasons for action which appear to exist cannot be accommodated within a purely agent-neutral system —or even perhaps within a general but agent-relative system —then we may have to modify our realist idea of value and practical reason accordingly. I don't mean to suggest that there is no conflict here. The opposition between objective reasons and subjective inclinations may be severe, and may require us to change our lives. I mean only that the truth, if there is any, will be arrived at by the exploration of this conflict rather than by the automatic victory of the most transcendent standpoint. In the conduct of life, of all places, the rivalry between the view from within and the view from without must be taken seriously.

III. ETHICS

1. In this lecture I want to take up some of the problems that must be faced by any defender of the objectivity of ethics who wishes to make sense of the actual complexity of the subject. There will be some parallels between what I say here and what I said in the first lecture, about the interpretation of objectivity with regard to the mind. Here also the treatment will be rather general and very incomplete. Essentially I shall discuss some examples in order to give grounds for believing that the enterprise is not hopeless.

In the second lecture I distinguished between agent-relative and agent-neutral values. Agent-neutral values, if there are any, are the values of things good or bad in themselves, things that there is reason for anyone to want or not to want. Agent-relative values, on the other hand, while they are also general, are defined *relatively*. They are specified by reference to the agent for whom they provide reasons. For example, if there were a reason for everyone to want the world to be a happier place, independently of the effect of this on him, that would be an *agent-neatral* value. If on the other hand each person had reason to want only his own happiness and the happiness of others whom he cared for, that would be an *agent-relative* value.

This contrast is central to an important set of issues about moral objectivity and its limits. Certain ethical positions, those sometimes called consequentialist, admit only agent-neutral values. That is, they hold that ethics is concerned only with what should *happen,* and never independently with what people should *do*. But the hegemony of agent-neutral values is challenged by two broad types of reasons that appear to be *agent-relative* in form, and

whose existence seems to be independent of agent-neutral values. It is these that I propose to discuss today.

The first type of reason stems from the desires, projects, commitments, and personal ties of the individual agent, all of which give him reasons to act in the pursuit of ends that are his own. These I shall collect under the general heading of reasons of *autonomy*.

The second type of reason stems from the claims of other persons not to be maltreated in certain ways. What I have in mind are not agent-neutral reasons for everyone to want it to be *the case* that no one is maltreated, but agent-relative reasons for each individual not to *maltreat others himself,* in his dealings with them (e.g., by violating their rights, breaking his promises to them, etc.). These I shall collect under the general, ugly, and familiar heading of *deontology.* Autonomous reasons would limit what we are *obliged* to do in the service of agent-neutral values. Deontological reasons would limit what we are *permitted* to do in the service of *either* agent-neutral *or* autonomous ones.

I am not sure whether all these agent-relative reasons actually exist. The autonomous ones are fairly intelligible; but while the idea behind the deontological ones can, I think, be explained, it is an explanation which throws some doubt on their validity. The only way to find out what limits there are to what we may or must do in the service of agent-neutral values is to see what sense can be made of the apparent limits, and to accept or reject them according to whether the maximum sense is good enough.

Taken together, autonomous, agent-neutral, and deontological reasons cover much of the territory of unreflective bourgeois morality. Common sense suggests that each of us should live his own life (autonomy), have some significant concern for the general good (agent-neutral values), and treat the people he deals with decently (deontology). It also suggests that these aims may produce serious inner conflict. Common sense doesn't have the last word in ethics or anywhere else, but it should be examined

before it is discarded.

Attempts have been made to find room for some version of both these types of apparent exception to agent-neutral ethics in a more complex agent-neutral system, using developments of con-sequentialism like rule-utilitarianism and motive-utilitarianism. I shall not try to show that these reductions of the agent-relative to the agent-neutral fail. Instead I shall present an alternative account of how the exceptions might make sense independently. The account in both cases depends on certain discrepancies between what can be *valued* from an objective standpoint, and what can be *seen* from an objective standpoint to *have value* from a *less* objective standpoint.

2. Let me begin with autonomy.

Not all the sources of subjective reasons are as simple as sensory pleasure and pain. These simply *evoke* an awareness of their value, without thought, choice, or deliberation, and I argued earlier that this makes it reasonable to affirm their value directly from an objective standpoint by judging that they are impersonally good or bad. Difficult as it may be to carry out, each of us has reason to give significant weight to the simple sensory pleasure or pain of others as well as to his own. I believe that when these values occur in isolation, the results can be rather demanding. If you and a stranger have both been injured, for example, and you have one dose of painkiller, and his pain is much more severe than yours, you should give him the painkiller. Not for any com-plicated reasons, but simply because of the relative severity of the two pains, which provides an agent-neutral reason to prefer the relief of the more severe. The same may be said of other basic elements of human good and evil.

But most human values are not like this. Though some human interests give rise to agent-neutral values (and not only pleasure and pain) I now want to argue that not all of them do. If I have a bad headache, anyone has a reason to want it to stop. But if for

instance I badly want to become a first-rate pianist, not everyone has a reason to want me to practice. *I* have a reason to want to practice, and it may be just as strong as my reason for wanting my headache to go away. But other people have very little reason, if any, to care whether I become a first-rate pianist or not. Why is this?

I think it is easier to believe in this distinction than to explain it.

There are two ways in which a value may be conditional on a desire: the value may lie either outside or inside the conditional, so to speak. In the former case, a person's having X if he desires X has agent-neutral value: satisfaction of the desire has objective utility that everyone has reason to promote. In the latter case, if a person desires X, his having X has agent-relative value for him: 'having' the value is conditional on having the desire, and satisfaction of the desire does not have agent-neutral utility.

Roughly, (and I really mean roughly) I think involuntary desires belong in the first category and desires that are *adopted* belong in the second. Most of the things we pursue, if not most of the things we avoid, are things we choose. Their value to us depends on our individual aims, projects, and concerns, including particular concerns for other people that reflect our relations with them, and they acquire value only because of the interest we develop in them and the place this gives them in our lives.

When we look at such desires objectively, from outside, we can acknowledge the validity of the reasons they give for action, without judging that there is an agent-neutral reason for any of those things *to be done*. That is because, when we move to the objective standpoint, we are not *occupying* the perspective from which these values have to be accepted. Their diversity and their dependence on the history and circumstances of the agent insures this. From a point of view outside the perspective of my ambition to become a first-rate pianist, it is possible to *recognize* and *understand* that perspective and so to *acknowledge* the reasons that arise

inside it; but it is not possible to *accept* those reasons as one's own, unless one adopts the perspective rather than merely recognizing it.

So objective *understanding* of such reasons does not imply objective *acceptance* of them; because in order to have and act on them one must occupy the perspective of a particular life and its aims. Whether you are subject to their value depends on what your values *are.* A bad headache, on the other hand, can be recognized as bad independently of the *values* of the person whose headache it is: it has nothing to do with his personal values. So you do not have to be him to have reason to want it to go away. Its badness is agent-neutral, and you have a reason to want it to stop whatever your values are.

There is nothing incoherent in wanting to be able to play all the Beethoven piano sonatas by heart, while thinking that impersonally it doesn't matter whether one can do this. In fact one would have to be deranged to think it *did* matter impersonally. It doesn't even matter much impersonally that *if* someone wants to play all the Beethoven sonatas by heart, he should be able to. It matters a little, so that if he is incapable of achieving it, it would be better if he didn't want to. This is because the realization of a personal ambition is pleasant and its frustration is painful, so the agent-neutral values of pleasure and pain come into effect here. But even that is a rather weak agent-neutral value, since it is not the impersonal correlate of the agent-relative reasons deriving directly from the ambition, whose object is not pleasure. If an interest is developed by the agent himself through his choices and actions, then the objective reasons it provides are primarily agent-relative. Any agent-neutral reasons stemming from it must express values that are independent of the particular perspective and system of preferences of the agent.

The general values of pleasure and pain, satisfaction and frustration, fill this role to some extent, as I have said, though only to the extent that they can be detached from the value of the

object of desire whose acquisition or loss produces the feeling. This, incidentally, explains the appeal of hedonism to consequentialists: it reduces all value to the impersonal common denominator of pleasure and pain.

But what there is *not,* I believe, is a general agent-neutral value of the satisfaction of desires and preferences. The strength of an individual's personal preferences in general determines what they give *him* reason to do, but they do not determine the agent-neutral value of his getting what he wants. That is because their satisfaction *has* value only from the standpoint of the *values expressed in* those preferences. There is no independent value of preference-satisfaction *per se* which preserves its force even from an impersonal standpoint.

This rather harsh position can be modified somewhat by admitting that there is another, more general, level at which agent-neutral values do appear when one considers the area of personal preferences objectively. That is the level of the background of choice, liberty, and opportunity which makes the development and pursuit of voluntary concerns possible. Someone's having the freedom and the means in a general way to lead his life is not a good that can be appreciated only through the point of view of the particular set of concerns and projects he has formed. It is a quite general good, like the goods of health, food, physical comfort, and life itself, and if agent-neutral value is going to be admitted at all, it will naturally attach to this. People have reason to care about the liberty and general opportunities of others as they have reason to care about their physical comfort. This is not equivalent to assigning agent-neutral value to each person's getting whatever he wants.

If this hypothesis of two levels of objectification is correct, then there is not a significant reason for something to happen corresponding to every reason for someone to do something. Each person has reasons stemming from the perspective of his own life which, though they can be publicly recognized, do not in general

provide reasons for others and do not correspond to reasons that the interests of others provide for him. He has some agent-neutral reasons to consider the interests of others, but the personal claims of autonomy, if they have significant weight, will allow him to pursue his own aims to some extent at the expense of those others. Since agent-relative reasons are general and not purely subjective, he must acknowledge that the same is true of others with respect to him.

All this is based on the assumption that an agent-neutral objecti-fication of this large category of individualistic subjective reasons does not make sense. But of course that doesn't entail that an *agent-relative* objectification is correct, instead. There is a radical alternative: it could be that these reasons have no objective valid-ity at all, agent-relative or agent-neutral. That is, it might be said by some utilitarian extremist that if there isn't an *agent-neutral* reason for me to learn the Beethoven sonatas by heart — if it wouldn't be a good thing in itself; if the world wouldn't be a better place for my being able to play all the Beethoven sonatas — then I have no reason of *any* kind to learn them, and I had better get rid of my desire to do so as soon as possible.

That is a logically possible move, but not, I think, a plausible one. It results from the aim of eliminating perspective from the domain of real value to the greatest possible extent, and that aim is not based on anything, so far as I can see. We should certainly try to harmonize our lives to some extent with how we think the world should be. But there is no necessity, I believe, to abandon all values that do not correspond to anything desirable from an impersonal standpoint, even though this may be possible as a personal *choice* — a choice of self-transcendence.

If there are, objectively, both agent-relative and agent-neutral reasons, this raises a problem about how life is to be organized so that both can be given their due. Just to offer a footnote about the relation between ethics and political theory, one way of dealing with this problem is to put much of the responsibility for securing

agent-neutral values into the hands of an impersonal institution like the state. A well designed set of political and social institutions should function as a moral buffer to protect personal life against the ravenous claims of impersonal good, and vice versa.

3. Let me turn now to the obscure topic of deontological constraints. These are agent-relative reasons which depend not on the aims or projects of the agent but on the claims of others. Unlike autonomous reasons, they cannot be given up at will. If they exist, they restrict what we may do in the service of either agent-relative or agent-neutral goals.

Whatever their explanation, they are conspicuous among the moral appearances. Here is an example to focus your intuitions.

You have an auto accident one winter night on a lonely road. The other passengers are badly injured, the car is out of commission, and the road is deserted, so you run along it till you find an isolated house. The house turns out to be occupied by an old woman who is looking after her small grandchild. There is no phone, but there is a car in the garage, and you ask desperately to borrow it and explain the situation. She doesn't believe you. Terrified by your desperation, she runs upstairs and locks herself in the bathroom, leaving you alone with the child. You pound ineffectively on the door and search without success for the car keys. Then it occurs to you that she might be persuaded to tell you where they are if you were to twist the child's arm outside the bathroom door. Should you do it?

It is difficult not to see this as a serious dilemma, even though the child's getting his arm twisted is a minor evil compared with your friends' not getting to a hospital. The dilemma must be due to a special reason against *doing* such a thing. Otherwise it would be *obvious* that you should choose the lesser evil, and twist the child's arm.

Common moral intuition recognizes several types of deontological reasons — limits on what one may do to people or how

one may treat them. There are the special obligations created by promises and agreements; the restrictions against lying; the prohibitions against violating various individual rights, rights not to be killed, injured, imprisoned, threatened, tortured, coerced, robbed; the restrictions against imposing certain sacrifices on someone simply as means to an end; and perhaps the special claim of immediacy, which makes distress at a distance so different from distress in the same room. There may also be a deontological requirement of fairness, of evenhandedness or equality in one's *treatment* of people. (This is to be distinguished from any *agent-neutral* value thought to attach to equality in the distribution of benefits, considered as an aspect of the assessment of *states of affairs.*)

In all these cases it appears that the special reasons, if they exist, cannot be explained simply in terms of *agent-neutral* values, because the particular relation of the agent to the outcome is essential. Deontological constraints may be *overridden* by agent-neutral reasons of sufficient strength, but they are not themselves to be understood as the expression of agent-neutral values of any kind. It is clear from the way such reasons work that they cannot be explained by the hypothesis that the violation of a deontological constraint has high negative agent-neutral value. Deontological reasons have their full force against *your doing* something —not just against its *happening.*

For example, if there really are such constraints, the following things seem to be true. It seems that *you* shouldn't break a promise or tell a lie for the sake of some benefit, even though you would not be required to forego a comparable benefit in order to prevent someone else from breaking a promise or telling a lie. And it seems that you shouldn't twist the arm of a small child to get its grandmother to do something, even if the thing is quite important — important enough so that it would not be reasonable to forego a comparable benefit in order to prevent someone else from twisting a child's arm. And it may be that you shouldn't

engage in certain kinds of unfair discriminatory treatment (in an official role, for example) even to produce a good result which it would be unreasonable to forego in order to prevent similar unfairness by *others.*

Some may simply deny the plausibility of such moral intuitions. Others may say that their plausibility can be subtly accounted for in terms of agent-neutral values, and that they appear to involve a fundamentally different type of reason for action only if they are inadequately analyzed. As I have said, I don't want to take up these alternative accounts here. They seem to me essentially *revisionist,* and even if from that point of view they contain a good deal of truth, they do not shed light on the deontological conceptions they are intended to replace. Sometimes, particularly when institutions and general practices are involved in the case, there may be an agent-neutral justification for what looks initially like an agent-relative restriction on action. But I am convinced there are many cases that evoke a different type of moral intuition. Right or wrong, it is this type of view that I want to explore and understand. There is no point in trying to show in advance that the controversy does not exist.

One reason for the resistance to deontological constraints is that they are *formally* puzzling, in a way that the other reasons we have discussed are not. We can understand how autonomous agent-relative reasons might derive from the specific projects and concerns of the agent, and we can understand how agent-neutral reasons might derive from the interests of others, giving each of us reason to take them into account. But how can there be *agent-relative* reasons to respect the claims of *others?* How can there be a reason not to twist someone's arm which is not equally a reason to prevent his arm from being twisted by someone else?

The agent-relative character of the reason cannot come simply from the character of the interest that is being respected, for that alone would justify only an agent-neutral reason to protect the interest. And the agent-relative reason does not come from an aim

or project of the individual agent, for it is not conditional on what the agent wants. Deontological restrictions, if they exist, apply to everyone: they are mandatory and may not be given up like personal ambitions or commitments.

There is no doubt that ideas of this kind form an important part of common moral phenomenology. Yet it is tempting to think that the whole thing is a kind of moral illusion resulting either from innate psychological dispositions or from crude but useful moral indoctrination. But this hypothesis faces problems in explaining what the illusion *is*. It may be a good thing if people have a deep inhibition against torturing children even for very strong reasons, and the same might be said of other deontological constraints. But that does not explain why we cannot come to *regard* it as a mere inhibition which it is good to have. An illusion involves a judgment or a disposition to judge, and not a mere motivational impulse. The phenomenological fact that has to be accounted for is that we seem to apprehend in each individual case an extremely powerful agent-relative *reason* not to torture a child. This presents itself as the apprehension of a *truth,* not just as a psychological inhibition. And the claim that such an inhibition is in general very useful does nothing to justify or explain the conviction of a strong reason in every individual case. That conviction is what has to be analyzed and accounted for, and accepted or rejected according to whether the account gives it an adequate justification.

4. I believe that the traditional principle of double effect, despite problems of application, provides a rough guide to the extension and character of deontological constraints, and that even after the volumes that have been written on the subject in recent years, this remains the right point of convergence for efforts to capture our intuitions.[6] The principle says that to violate

---

[6] A good statement of a view of this type is found in Charles Fried's recent book, *Right and Wrong* (Cambridge: Harvard University Press, 1978).

deontological constraints one must maltreat someone else *intentionally.* The maltreatment must be something that one does or chooses, either as an end or as a means, rather than something one's actions merely cause or fail to prevent, but that one doesn't aim at.

It is also possible to *foresee* that one's actions will cause or fail to prevent a harm that one does not *intend* to bring about or permit. In that case it is not, in the relevant sense, something one *does,* and does not come under a deontological constraint, though it may still be objectionable for impersonal reasons. (One point worth stressing: the constraints apply to intentionally *permitting* as well as to intentionally *doing* harm. Thus in our example, there would be the same kind of objection if *with the same end in view* you permitted someone else to twist the child's arm. You would have let it happen *intentionally,* and that would be different from a failure to prevent such an occurrence because you were too engaged in doing something *else* which was more important.)

So far this is just moral phenomenology: it does not remove the paradox. *Why* should we consider ourselves far more responsible for what we do (or permit) intentionally than for consequences of action that we foresee and decide to accept but that do not form part of our aims (intermediate or final) ? How can the connection of ends and means conduct responsibility so much more effectively than the connection of foresight and avoidability ?

It is as if each action produced a special perspective on the world, determined by intention. When I twist the child's arm intentionally I incorporate that evil into what I do: it is my creation and the reasons stemming from it are magnified from my point of view so that they tower over reasons stemming from greater evils that are more 'distant' because they do not fall within the range of intention.

That is the picture, but how can it be correct ?

I believe that this is one of those cases in which the removal of paradox is not a philosophical advance. Deontological reasons

are essentially problematic, and the problem is an instance of the collision between subjective and objective points of view. The issue is whether the special, personal perspective of agency has fundamental significance in determining what people have reason to do. The question is whether, because of this perspective, I can have sufficient reason not to do something which, considered from an external standpoint, it would be better if I did. That is, *things* would be better, what *happened* would be better, if I twisted the child's arm than if I did not. But I would have *done* something worse. If considerations of what I may do, and the correlative claims of my victim, can outweigh the substantial impersonal value of what will happen, that can only be because the perspective of the agent has an importance in practical reasoning that resists domination by a conception of the world as a place where good and bad things happen, and have their value without perspective.

I have already claimed that the dominance of this agent-neutral conception of value is not complete. It does not swallow up or overwhelm the agent-relative reasons arising from those individual ambitions, commitments, and attachments that are in some sense chosen. But the admission of what I have called autonomous agent-relative reasons does not imply the possibility of deontological reasons. The two are very different. The special paradox of deontological reasons is that although they are agent-relative, they do not express the subjective autonomy of the agent at all. They are *demands*. The paradox is that this partial, perspectival respect for the interests of others should not give way to an impersonal respect free of perspective. The deontological perspective seems primitive, even superstitious, by comparison: merely a stage on the way to full objectivity. How can what we *do* in this narrow sense be so important?

5. Let me try to say where the strength of the deontological view lies. We may begin by considering a curious feature of

deontological reasons on which I have not yet remarked. Intention appears to magnify the importance of *evil* aims by comparison with *evil* side effects in a way that it does not magnify the importance of *good* aims by comparison with *good* side effects. We are supposed to avoid using evil means to produce a good end, even though it would be permissible to produce that good end by neutral means with comparably evil side effects.

On the other hand, given two routes to a legitimate end, one of which involves *good* means and *neutral* side effects, and the other of which involves *neutral* means and slightly *better* side effects, there is no reason to choose the first route. Deontological reasons tell us only *not* to aim at *evil;* they don't tell us *to* aim at *good,* as a means. Why should this be? What is the relation between evil and *intention,* or *aiming,* that makes them clash in a special and intense way?

The answer emerges if we ask ourselves what is the essence of *aiming,* what differentiates it from merely producing a result knowingly?

The difference is that action intentionally aimed at a goal is *guided* by that goal. Whether the goal is an end in itself or only a means, action aimed at it must follow it and be prepared to adjust its pursuit if deflected by altered circumstances. Whereas an act that merely *produces* an effect does not *follow* it, is not *guided* by it, even if the effect is foreseen.

What does this mean? It means that to aim at evil, *even as a means,* is to have one's action *guided* by evil. One must be prepared to adjust it to insure the production of evil: a falling off in the level of the desired evil must be grounds for altering what one does so that the evil is restored and maintained. But the *essence* of evil is that it should *repel* us. If something is evil, our actions should be guided, if they are guided by it at all, toward its elimination rather than toward its maintenance. That is what evil *means.* So when we aim at evil we are swimming head-on against the normative current. Our action is guided by the goal at every

point in the direction diametrically opposite to that in which the value of that goal points. To put it another way, if we aim at evil we make what we do in the first instance a positive rather than a negative function of evil. At every point, the intentional function is simply the normative function reversed, and from the point of view of the agent, this produces the acute sense of doing something awful.

If you twist the child's arm, in our example, your aim is to produce pain. So when the child cries, "Stop, it hurts!" his objection corresponds in perfect diametrical opposition to your intention. What he is pleading as your reason to stop is precisely your reason to go on. If it didn't hurt, you would twist harder, or try the other arm. You are pushing directly and essentially against the normative force intrinsic to your goal, for it is the production of pain that guides you. It seems to me that this is the essence of deontological constraints. What feels peculiarly wrong about doing evil intentionally even that good may come of it is the headlong striving against value that is internal to one's aim.

Some corroboration of this diagnosis may be found by asking what *would* be the corresponding principle governing the relation between intention and good, as opposed to evil? I have said that there is no deontological requirement to aim at good — only a requirement not to aim at evil. But the analogue of the requirement not to aim *at* evil would be a requirement not to aim *away* from good. To aim to *prevent* something good as a means to a worthy end would have a similar quality of normative reversal, though less acute than that of aiming at evil. And I believe there may be deontological constraints, though not such conspicuous ones, against deliberately preventing something good, in order that good may come of it. (Think for example of someone who resists ameliorating the condition of the poor because he thinks it will reduce their anger and diminish the long-term chance of a social revolution.)   I mention the point, but will not pursue it.

But all this still leaves unsettled the question of justification.

For it will be objected that if one aims at evil as a means only, then one's action is not really being guided by evil but by overall good, which includes a *balance* of goods and evils. So when you twist the child's arm, you are guided by the aim of rescuing your injured friends, and the good of that aim dominates the evil of the child's pain. The immediacy of the fact that you must try to produce evil as a subsidiary aim is phenomenologically important, but why should it be morally important?

Here I think we have come down to a fundamental clash between perspectives. The question is whether to disregard the resistance encountered by my immediate pursuit of evil, in favor of the overall value of the results of what I do. When I view my act from outside, and think of it as resulting from a choice of the impersonally considered state of the world in which it occurs, this seems rational. In thinking of the matter this way, I abstract my will and its choices from my person, as it were, and even from my actions, and decide directly among states of the world, as if I were taking a multiple-choice test. If the choice is determined by what on balance is impersonally best, then I am guided by good and not by evil.

But the self that is so guided is the objective self which regards the world impersonally, as a place containing TN and his actions, among other things. It is detached from the perspective of TN: for it views the world from nowhere within it. It chooses, and then TN, its instrument, or perhaps one could say its agent, carries out the instructions as best he can. *He* may have to aim at evil, for the impersonally best alternative may involve the production of good ends by evil means. But he is merely following orders.

To see the matter in this light is to see both why the appeal of agent-neutral, consequentialist ethics is so great and why the contrary force of agent-relative, deontological ethics is so powerful The detached, objective view takes in everything and provides a standpoint of choice from which all choosers can agree about what should happen. But each of us is not only an objective self but a

particular person with a particular perspective; we act in the world with that perspective, and not only from the point of view of a detached will, selecting and rejecting world-states. So our choices are not merely choices of states of the world, but of *actions.* From this point of view, the pursuit of evil in twisting the child's arm looms large. The production of pain is the immediate aim, and the fact that from an external perspective you are choosing a balance of good over evil does not cover up the fact that this is the kind of action you are undertaking.

This account of the force of deontological reasons applies with special clarity to the constraint against doing harm as a means to your ends. A fuller deontological theory would have to explain the different types of normative grain against which one acts in breaking promises, lying, discriminating unfairly, and denying immediate emergency aid. It would also have to deal with problems about what exactly is being aimed at in cases of action that can be described in several different ways. But I believe that the key to understanding any of these moral intuitions is the distinction between the internal viewpoint of the agent in acting and an external, objective viewpoint which the agent can also adopt. Reasons for action look different from the first point of view than from the second.

So we are faced with a choice. For the purposes of ethics, should we identify with the detached, impersonal will that chooses world-states, and act on reasons that are determined accordingly? Or is this an evasion of the full truth about who we really are and what we are doing, and an avoidance of the full range of reasons that apply to creatures like us? If both personal and impersonal perspectives are essential to us, then it is no wonder that the reasons for action deriving from them do not fit comfortably together.

6. I believe this is a true philosophical dilemma which has no natural resolution. It arises out of our nature, which includes

different points of view on the world. In ethics the contest between objective detachment and the individual perspective is acute. We feel it in the dilemma between deciding on the basis of action and deciding on the basis of outcome; in the dilemma between living one's private life and serving the general good; in the dilemma between concern for what is actually happening and concern for what is timeless. When we ask ourselves how to live, the complexity of what we are makes a unified answer difficult.

It is conceivable that this may change, for we are at a very primitive stage of moral development. Even the most civilized human beings have only a haphazard understanding of how to live, how to treat others, how to organize their societies. The idea that the basic principles of morality are *known,* and that the problems all come in their interpretation and application, is one of the most fantastic conceits to which our conceited species has been drawn. (The idea that, if we cannot easily know it, there is no truth here, is no less conceited.) Not all of our ignorance in these areas is ethical, but a lot of it is. And the idea of the possibility of moral progress is an essential condition of moral progress. None of it is inevitable.

It would be foolish to try to lay down in advance the outlines of a correct method for ethical progress; but I believe that the general direction that it is reasonable to follow *at present* is connected to the awkward pursuit of objectivity that we have been discussing. This does not mean that greater detachment always takes us closer to the truth. Sometimes, to be sure, objectivity will lead us to regard our original inclinations as mistaken, and then we will try to replace them or bracket them as ineliminable but illusory. But it would be a mistake to try to eliminate perspective from our conception of ethics entirely —— as much of a mistake as it would be to try to eliminate perspective from the universe. Though it may be equally tempting, it would be no more reasonable to eliminate all those reasons for action that cannot be assimilated to the most objective, impersonal system of value than it

would be to eliminate all facts that cannot be assimilated to physics.

At the same time, I believe that the agent-relative values of individual autonomy and of deontology are severely challenged when we look at our lives from outside.

One sign of this is that the most common form of subjective resistance to objective dominance is just *badness*. I have not discussed this familiar phenomenon so far because while it is part of the struggle between subjective and objective standpoints in practical reason, it has to do neither with the content nor with the reality of ethics. To be bad is to recognize the claims of morality from an objective standpoint but to refuse to submit to them. One may not admit that this is what one is doing — may offer fake justifications or rationalizations — but recognition of the general principles that one is defying can show up in other ways: in one's reaction when subjected to comparable treatment by others, for example. Badness is not the same as amorality. On the contrary, it shows that one *accepts* the reality and objectivity of ethics. It is the most direct form of subjective resistance to that objective standpoint that forms a part of each of us, and whose demands can be so exhausting.

Since the subjective–objective struggle can take this form, there is room for considerable self-deception. It is not always easy to tell, for example, whether a morality that leaves extensive free space in each individual life for the pursuit of personal interests is not just a disguise for the simplest form of badness: selfishness in the face of the legitimate claims of others. It is hard to be good, as we all know.

I suspect that if we try to develop a system of reasons which harmonizes personal and impersonal claims, then even if it is acknowledged that each of us must live in part from his own point of view, there will be a tendency for the personal components to be altered. As the claims of objectivity are recognized, they may come to form a larger and larger part of each individual's con-

ception of himself, and will influence the range of personal aims
and ambitions, and the ideas of his particular relations to others.
I do not think it is utopian to look forward to the gradual develop-
ment of a greater universality of moral respect, an internalization
of moral objectivity analogous to the gradual internalization of
scientific progress that seems to be a feature of modern culture.

On the other hand, there is no reason to expect progress to be
reductive, though here as elsewhere progress is too easily identi-
fied with reduction and simplification. Distinct individuals are
still the clients of ethics, and their variety guarantees that plu-
ralism will be an essential aspect of any adequate morality, how-
ever advanced.

There have to be principles of practical reason that allow us
to take into account values that we do not share, but whose force
for others we must acknowledge. In general, the problem of how
to combine the enormous and disparate wealth of reasons that
practical objectivity generates, together with the subjective reasons
that remain, by a method that will allow us to act and choose in
the world, is dauntingly difficult.

And this brings us to a final point. There can be no ethics
without politics. A theory of how individuals should act requires
a theory — an ethical theory, not just an empirical one — of the
institutions under which they should live: institutions which sub-
stantially determine their starting points, the choices they can
make, the consequences of what they do, and their relations to
one another. Since the standpoint of political theory is necessarily
objective and detached, it offers strong temptations to simplify,
which it is important to resist. A society must in some sense be
organized in accordance with a single set of principles, even
though people are very different.

This is inconvenient: it may seem that political theory must be
based on a universal human nature, and that if we cannot dis-
cover such a thing we have to invent it, for political theory must
exist. To avoid such folly, it is necessary to take on the much

more difficult task of devising fair uniform social principles for beings whose nature is not uniform and whose values are legitimately diverse. If they were diverse enough, the task might be impossible — there may be no such thing as intergalactic political theory — but within the human species the variation seems to fall within bounds that do not rule out the possibility of at least a partial solution. This would have to be something acceptable from a standpoint external to that of each particular individual, which at the same time acknowledges the plurality of values and reasons arising within all those perspectives. Even though the morality of politics is rightly more agent-neutral than the morality of private life, the acknowledgment of agent-relative values and autonomy is essential even at the level that requires the greatest impersonality.

There is no telling what kinds of transcendence of individuality will result over the long term from the combined influence of moral and political progress, or decline. At the moment, however, a general takeover of individual life from the perspective of the universe, or even from the perspective of humanity, seems premature — even if some saints or mystics can manage it. Reasons for action have to be reasons for individuals, and individual perspectives can be expected to retain their moral importance so long as diverse human individuals continue to exist.