Morality and Consequences

JONATHAN BENNETT

THE TANNER LECTURES ON HUMAN VALUES

Delivered at
Brasenose College, Oxford University

May 9, 16, and 23, 1980

Jonathan Bennett was born in New Zealand in 1930 and educated there and at the University of Oxford. He taught at Cambridge for twelve years, for eleven more in Canada, and since 1979 has been Professor of Philosophy at Syracuse University. He has written a good many articles— including, on moral philosophy, 'Whatever the Consequences' and 'The Conscience of Huckleberry Finn' — two books on the philosophy of mind and language, and three on early modern philosophy. He collaborated with Peter Remnant in a recently published edition and translation of Leibniz's *New Essays,* and is currently writing a commentary on Spinoza's *Ethics* and a book on moral philosophy which is closely related to his lectures in this volume.

I. KILLING AND LETTING DIE

I want to express my gratitude to the Principal of Brasenose and Mrs Nicholas, and to the Fellows of Brasenose, for the warming hospitality they have extended to my wife and myself; to Professor Tanner for his magnificent benefaction; and to all of you for being here. For me this time in Oxford — my first solid visit since I graduated from here exactly twenty-five years ago — is a heavily charged occasion. Added to the pride and anxiety which go with being the Tanner Lecturer, there is the joy of simply being in Oxford, and the complex set of emotions — known collectively as nostalgia — which are stirred by looking back across half a lifetime. In my case those emotions are strongly coloured by the fact that of my Oxford contemporaries the four who were dearest to me are all dead — have all been dead for many years now. One of the things I am doing in returning here is to celebrate the memories of Donald Anderson, Robin Farquharson, John Lemmon, and Richard Selig.

<div align="center">∗ ∗ ∗</div>

In this lecture I shall offer to make clear, deeply grounded, objective sense of a certain contrast: I call it the contrast between positive and negative instrumentality, and it shows up in ordinary speech in remarks about what happens because a person *did* do such and such, as against what happens because he *did not.*

The line between positive and negative instrumentality lies fairly close to some others which are drawn by more ordinary bits of English. For instance, the difference between positive and negative instrumentality in someone's dying is cousin to the difference between killing a person and letting him die. The latter distinction has the advantage of being already encoded in plain

untechnical English; but it also has drawbacks for the sort of
moral philosophy I want to do, as I shall now explain.

First, I want a genuine distinction — something which marks
off two mutually exclusive species; and my second desideratum
is that the two be jointly exhaustive of a genus which I call that
of 'prima facie responsibility' for a state of affairs. I want it to
include every case where a person's conduct makes him in some
way and to some degree responsible for a given state of affairs.
This is to be decidable in advance of considering whether he
should be excused on grounds of mental incompetence, unavoid-
able ignorance, or whatever. Just because those matters are so
morally important, I want them to have their own separate day
in court; so I don't want the line I am drawing to get tangled
up with them anywhere along its length.

Third, because the distinction is to separate out two classes of
situation so that we can do some basic moral thinking about them,
it must not depend for its initial application on our having already
done some of the moral thinking. So it must be defined in terms
which have no moral content in their meanings: if they turn out
to have moral import, that will emerge later as a matter of sub-
stantive moral judgment: it will not be there all along as a matter
of meaning.

Fourth, the line to be drawn should be statable in terms which
are clear, objective, and deeply grounded in the natures of things.
I do not want it to be one whose application to particular cases is
at the mercy of controversy; or even at the mercy of agreed lin-
guistic intuitions if these are not backed by a decent degree of
clarity about what they are intuitions *of*. That is a matter of
degree and is vague, but it will get a little clearer as I go along.

Those desiderata are better satisfied by the line between posi-
tive and negative instrumentality — the difference between 'be-
cause he did' and 'because he didn't' — than by any other distinc-
tion which might be regarded as a rival to it. I shall mainly dis-
cuss one rival, namely the line which has causal verbs on one side

of it and corresponding phrases about 'letting' things happen on
the other side — felling and letting fall, misleading and letting
go astray, spoiling and letting deteriorate, killing and letting die.
This line is worse for my purposes than the line between positive
and negative instrumentality because it satisfies none of my four
desiderata.

First, it separates two non-overlapping classes of verbal expres-
sion, but not two non-overlapping classes of event. There are
killings which get described as lettings die (such as pulling the
plug on the life-support system of a terminal patient), and there
are lettings die which get described as killings (such as killing a
houseplant by not watering it).

Second, the two are not jointly exhaustive of the genus 'prima
facie responsibility'. There are cases where something happens
because I did not do A, but where, since I did not know that it was
liable to happen, it is improper to say that I 'let' it happen. If I
didn't know, then perhaps I am not morally accountable for its
happening; but that is a matter for subsequent moral discussion
which I don't want to be preempted by the very terms in which my
line is initially drawn. And, on the other side, there are cases
where something happens because I did do A but where the
relevant causal verb is not applicable — although she died because
of what I did, I didn't kill her but merely hired or forced some-
one else to kill her. Again, there are moral issues about the differ-
ence between that and outright killing; and again I want to set
those aside for later consideration rather than building them into
the initial distinction.

Third, along some of its length the line between doing and
letting happen — e.g., killing and letting die — reflects prior
moral judgments. For example, if a houseplant dies of drought,
and would have survived if I had watered it, the question of
whether I *killed* it depends largely upon whether it was my job,
my responsibility, to water it. That is the sort of moral input or
moral taint which I want to keep out of my basic distinction.

Fourth, and last, there is controversy about parts of the border-line around killing and letting die, and even where there is agreement, there is sometimes not enough clarity about what the underlying principles are. For instance, we speak of pulling the plug on someone's respirator as a case of 'letting' him die because we see his dying as something which is tending or trying or straining to happen, and we see what we are doing as the mere removal of an obstacle to that process. I cannot find that that way of viewing the situation corresponds to anything in the objective world which I would be prepared to make room for in my moral thinking. I might have to withdraw that remark: someone might reveal what lies behind those removal-of-an-obstacle intuitions, and show it to be fit to bear a heavy moral load. Until such a revelation comes along, however, I add this to my charge-list against the distinction between doing and letting happen.

There are similar drawbacks to most of the other terminology that is commonly used to mark distinctions which, since they partly coincide with the positive/negative line, could be regarded as rivals to it.

For example, the meanings of 'refrain' and 'forbear' are too restricted: either of these terms, when combined with any of its plausible partners, yields a distinction which is not exhaustive of the genus. If we take the line between 'because he did A' on the one side and 'because he refrained from doing A' on the other, we shall be excluding cases where he did not do A but did not *refrain* from doing it either, because it never entered his head to do it, or because it occurred to him to do it but he felt no inclination that way; and similarly with 'forbear'.

The situation with 'omit' is different, but no better. It seems that you can 'omit' to do something without feeling a pull towards doing it; but you can't properly be said to 'omit' to do something unless you prima facie ought to have done it; and so we have a substantial moral taint in the language of act/omission if the latter term is properly used.

When people contrast 'active' with 'passive' euthanasia, they may be pointing to the positive/negative line which interests me. If they are — and indeed even if they are not — they are using 'passive' in a manner which seems not to stand up to critical scrutiny.

Worst of all is the verb 'to cause'. There are idiomatically natural ways of using it to draw something close to the positive/ negative line. If something happens because I did do A, it will very often be natural to say that I *caused* it to happen; and if it happens because I did not do A, it will often be natural to say that I didn't cause but *allowed* it to happen. But I cannot turn this to account in theory-building, because I cannot see how to make these idioms put their feet firmly enough on the ground.

If we tie the word 'cause' to any of the most promising philosophical theories about causes — e.g., Mackie's about INUS conditions, or Lewis's counterfactual analysis — then it won't do anything like the work of positive instrumentality. For according to those theories, if the door slams because I do not grab it, my not grabbing it can easily qualify as a cause of its slamming.

Those theories, however, concern 'cause' as a noun or a verb used in relating one event to another — 'e causes f' or 'e is a cause of f'. When 'cause' is used to draw something like the positive/ negative line, it is being used as a verb with a *person* as subject — 'He causes the door to close'. These uses of 'cause' have, so far as I know, no plausible, strong, clear philosophical theory to back them up: we seem to have to steer pretty much by our intuitions. And if we are to be guided by nothing but the linguistic proprieties, we shall find that plenty of negative cases will still be cases of someone's 'causing' something to happen, so that there will again be overlaps between causing things to happen and, for instance, allowing them to happen; and, worse still from my point of view, the criteria for whether a given negative instrumentality is a causing or not are themselves partly moral.*

---

* Judith J. Thornson, *Acts and Other Events* (Ithaca, N.Y.: Cornell University Press, 1977), p. 215.

Even if there were not that moral taint, the reliance on largely unexplained linguistic intuitions is for me a large drawback to the use of the verb 'to cause' in basic moral philosophy. I mentioned this a few moments ago, but should say a little more about it. What I am rejecting is the idea of taking unexplained linguistic intuitions as components in the hard data of my moral theorizing. If someone proposes, as a basic moral principle, something to the effect that it is worse to cause the death of an innocent person than to allow such a death, I don't know how to think about this except insofar as I am clear about what the difference is between causing a death and allowing one. Some people, on the contrary, are prepared to accept such a principle in advance of being clear about where it will lead them; for them, the pursuit of clarity about causing is part of the process of moral discovery. It is presumably because there are such people that we find, in the morally oriented literature on causing, ordinary-language semantics intertwined with moralising: writers take their stand on who causes what, as a way of jockeying for moral position. I don't mean to sneer at this, and when I revert to it at the end of my third lecture I'll indicate one respectable basis for taking this approach to such matters. But it is not my approach. Rather than holding firm to a principle using the verb 'to cause', and exploring the verb's meaning in order to discover what I am morally committed to, I would regard any unclarity over what 'to cause' means as automatically limiting my commitment to any moral principle containing it.

As I said, I favour the contrast between positive and negative, between 'because he did' and 'because he didn't'. As a point of reference for discussion of this contrast, let me introduce three very short stories. In each, a vehicle is on ground sloping down to a cliff top; and in each, there is a course of events which culminates in the vehicle's falling down the cliff. My interest is in the role in the different stories of someone I call John.

   A. John gives the vehicle the push which starts it rolling, and then nothing can stop it.

    B. The vehicle is rolling when the story starts. There is a rock in its path which would stop it. John kicks away the rock.

    C. The vehicle is already rolling. There is a rock near its path which would, if interposed, stop it. John does not interpose the rock.

The line which interests me falls between A and B on the one side and C on the other. In both A and B, the vehicle is destroyed because John did do such and such, while in C it is destroyed because he did not do such and such. The line between doing and letting falls differently, because most people say that in B John, in removing the rock, 'lets' the vehicle go to its destruction as he does also in C, that being just the sort of thing I dislike about the verb 'to let'.

These days, moral philosophers with an interest in theoretic foundations shy away from the positive/negative distinction, apparently because they are nervous about the concept of a negative action. Although I have no need for that concept, I shall say a little about its prospects.

Whether there can be a coherent concept of 'negative action' depends on what one's underlying ontology of actions, and thus presumably of events generally, is like. If Kim and Goldman are right, then actions are abstract entities and can perfectly well be negative. Really, Kim's 'actions' are *facts* about agents: just as the fact that John does not interpose the rock is distinct from the fact that John keeps both his feet on the ground, so Kim will say that John's non-interposition of the rock is one action and his keeping of both his feet on the ground is another. And, just as facts can be negative, so can actions if they are the finely-sliced, abstract items that Kim makes them out to be.

If, on the other hand, Lemmon and Quine are right, actions are concrete chunks of space-time, so that the phrase 'John's non-interposition of the rock at time T' is just one name for the

totality of what John is up to at time T, this being an entity which may also answer to such descriptions as 'John's keeping both feet on the ground at T'. There is no chance of making that entity negative in itself: negativeness is always *de dicto,* not *de re;* but the totality of what John is up to at time T is a *res,* a concrete particular thing, and cannot be negative. It answers *to* some negative descriptions, but then so does everything.

Bentham's celebrated account of negative actions seems to have fallen foul of this point. The only way I can make sense of what he wrote is to suppose that he took actions to be concrete chunks of space–time while also thinking that a subclass of them are negative; so that to pick out the members of the subclass we must be able to peer at the totality of what a man is up to at a given time and declare *it* to be negative, negative in itself, negative *de re.* Someone with those ideas at the back of his mind will be apt to conclude — as apparently Bentham did — that such an item can be negative only if it consists in the extreme of inaction, i.e., in a state of affairs which brackets the agent with corpses and fence posts and pebbles: "Acts . . . may be distinguished ... into positive and negative. By positive are meant such as consist in motion or exertion: by negative, such as consist in keeping at rest; that is, forbearing to exert one's self in such and such circumstances." [*] Of course Bentham doesn't stick to this disastrous account. He says, for instance, that the non-payment of a debt is 'it negative nction', without asking whether the defaulter keeps stock-still at the time when he should be paying up. Still, there the official account sits — 'by negative, such as consist in keeping at rest'. I think it is the result of a doomed attempt to use 'is negative' as a monadic predicate which applies to actions understood as concrete particulars.

In my opinion we shall not have a worthwhile ontology of events and actions unless their identity conditions lie between

---

[*] Jeremy Berntham, *An Introduction to the Principles of Morals and Legislation,* ch. 8, sec. 8.

Kim's and Quine's, so that a given bit of space–time can contain more than one event (which is all Quine allows) without any risk of having to accommodate the infinity of events which Kim threatens to cram into it. Davidson seems to be hoping for an intermediate position, though he has never found one. If someone does, we shall have to look at the details before we can know whether it allows for negative events or actions.

Anyway, as I said, I dispense with negative actions. In situation C, it doesn't matter whether John performs a negative action; the important negative item in C is a *fact,* namely the fact that John does not interpose the rock.

There are plenty of negative facts about his conduct in situation B also — he does not dance a jig, does not join the army, and so forth — but in B there is no negative fact about his conduct which, combined with the impersonal facts of the situation, guarantees that the vehicle will go over the cliff. If I tell you about the rolling vehicle with the rock in its path, and add only some negative fact such as that John does not stand on his head, you cannot tell from that what will become of the vehicle. Whereas in C the negative fact that John does not interpose the rock is all we need, given the impersonal circumstances, to know that the vehicle is doomed. In each case the crucial question is: what is the weakest fact about John's conduct which suffices for the vehicle's destruction? In B it is a positive fact, while in C it is negative.

That indicates in a rough way how I distinguish 'because he did' from 'because he didn't', or positive from negative instrumentality. At least it shows how I use the concept of a negative fact and liberate myself from the concept of a negative action.

But that leaves me with the enormous problem of explaining what it is for a fact or a proposition to be negative. It is easy to classify sentences as negative or not; but that is no help to me, since the very same proposition may be expressible in sentences of both sorts — e.g., the proposition that George *is not* in the armed forces any longer, or *is* now a civilian. In short, granted that

negativeness won't fall from propositions down onto actions or events, can I prevent it from floating off propositions up onto sentences? If I can't, my concept of negative instrumentality will be revealed as language-dependent: what happens in situation C will involve John's negative instrumentality if we describe his conduct as his *not* interposing the rock, but it will come out as positive if we fake up some non-negative sentence which means the same. For instance, let us give the word 'pormit' a meaning such that 'x pormits y with z' means that x does not put z in the path of y; and then we can say that in situation C the vehicle is destroyed because 'John pormits it with the rock', which is not a negative sentence. Such a concept of negative instrumentality, being at the mercy of choice of wording, would not be worth having; and rather than persist with it I would prefer to go back and try to revive one of the other candidates.

But I don't capitulate. I cannot offer — and suspect that there cannot be — a general theory of proposition-negativeness. But I can produce a very limited one which is nevertheless broad enough to support an account of negative instrumentality. As well as being limited, it will be shallower than one might have hoped: its shallowest roots won't run deep enough to let the whole account belong to abstract logic or fundamental metaphysics; but they will get below the surface of the language and will give to the resultant line between kinds of instrumentality a good deal more depth and objectivity than I can find for any of its rivals.

The underlying idea has occurred to philosophers as different as Kant and Ayer (I refer to Kant's notion of an 'infinite' proposition and Ayer's 'Negation' paper in his *Philosophical Essays)*. It is that a negative proposition is just an extremely though not entirely contentless one — a proposition which says enormously much less than its contradictory. It is the idea that a negative proposition sprawls across nearly but not quite the whole of logical space, is true at nearly but not quite all possible worlds.

That idea, as it stands, is indefensible. But there is a limited

adaptation of it which does work, and which is all I need. To introduce it, I should first explain why the idea cannot be rescued in its full generality.

Let us represent logical space — or the totality of possible worlds — by a square, each point in which represents one possible world, one complete way things might be. Then any proposition P is represented by the set of points representing worlds at which P is true. Thus

<p style="text-align:center">Roses are red and violets are blue</p>

is represented by the set of all the points which represent complete descriptions of worlds at which roses are red and violets blue; and that set can be thought of as a subregion within the square. It in turn is part of the larger region corresponding to

<p style="text-align:center">Roses are red,</p>

and is indeed just the intersection between that and the region which represents *Violets are blue.* Within this way of representing things, it is quite clear what is going to count as a representation of one proposition's saying more — having more content — than another; it will be the former's having a smaller area than the other. If P entails Q and not vice versa, then in a good sense P has more content than Q; and they will be drawn as in Figure 1, which manifestly gives P less area than Q. The link between amount of content and smallness of area is also embodied in the
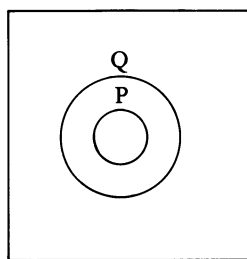

<p style="text-align:center">FIGURE 1</p>

fact that each *point* in the square represents a proposition which chatters on until it has said the whole truth about some possible world.

So far, so good: if one area contains and is not contained in another, you don't need a metric to tell you that the former is bigger than the latter — sheer topology is enough. But the Kant– Ayer idea about negative propositions requires a metric for logical space: we have to be able to compare the regions assigned to two propositions of which neither entails the other. Specifically, we need a basis for saying not merely that P and not-P divide up logical space between them, but further that the line falls not like that in Figure 2, but rather like that in Figure 3. No one has found a general procedure which will settle the question for each contradictory pair of propositions. It obviously cannot be done by counting possible worlds, for in every interesting case there are infinitely many of those — i.e., the same number of those — on each side of the line. I believe that there is no general solution to this problem, and that a fully general concept of negative proposition cannot be had through the Kant–Ayer idea — and therefore, I am pretty sure, cannot be had at all.

Suppose we set our sights a bit lower. Let us take a square which represents not the whole of logical space, but just the totality of possibilities with respect to one particular thing   at one particular time T. The points on this square represent propositions of the form   Fa at T
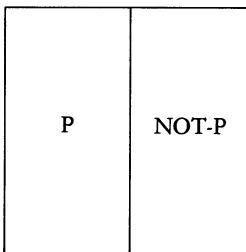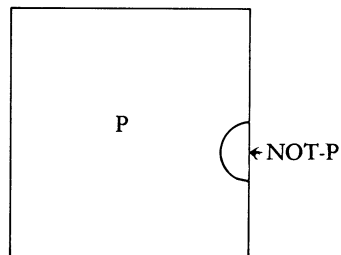

FIGURE 2   FIGURE 3

where 'F' stands for an absolutely specific statement about *a's* state at that moment. A line across that square represents a pair of propositions which are complementary within the square, though they are not strictly contradictories because each of them entails the existence of a at T. For example, *John's right hand moves at T* and *John's right hand does not move at T,* where the square represents the possible conditions of John's body, or perhaps of John's right hand.

It will help me to explain what I'm doing — though it isn't theoretically required — if I thin out the propositions represented on the square by restricting the values of F to predicates concerning motion. That is, each point proposition is an absolutely specific statement about whether and how the object a is moving at time T.

That doesn't immediately help to show how a line can fall unevenly across a square, for it still leaves us with equal numbers of point propositions on the two sides of the line. Suppose we have the square for John's body at T, and take the complementary pair *John's right hand moves at T* and *John's right hand does not move at T :* there are infinitely many ways of being in movement with your right hand moving, and infinitely many ways of being in movement with your right hand not moving; and so once again the count will be the same on each side of the line. This difficulty will always arise except where one member of the complementary pair is itself a point proposition or a finite disjunction of them; but such propositions are of no interest to moral philosophy. So I shall ignore them, attending only to complementary pairs each of whose members corresponds to a region of the square rather than to a finite set of points. And so, as I said, it is no use counting point-propositions as a way of getting the line between a complementary pair to divide the square unequally.

As a start towards showing what *can* make a line fall unequally across a square, I shall first show how two mutually contrary propositions can be entitled to equal-sized regions of a

square, i.e., can deserve to be represented as in Figure *4,* rather than, for instance, as in Figure 5, or in some manner which implies nothing metrical. I suggest that the former of those diagrams is the right one if P and Q differ only in that they attribute *different but equally specific* modes of motion to the object a. Thus, if P says that at time T *a* is drifting in the direction NE (plus or minus 1.5°) at a speed of 5 m.p.h. plus-or-minus 1 m.p.h), and Q says that at that time *a* is drifting SW at 7 m.p.h. (with the same plus-or-minus riders), then we have two equally specific attributions of movement to the same thing at the same time. I regard this as providing a good sense for the claim that the two propositions have the same amount of content and thus should be represented by equal-sized regions of the square.

It works with motion but won't work with everything. If, say, I had a square full of propositions about *a's* colour at time T, I could assign propositions equal areas only if I knew whether, for instance, 'shocking pink' and 'royal blue' are equally specific; but I don't know that, and we have no objective basis on which to decide it. But because we have objective measures for space and time, we have them also for movement, and thus for degree of specificity of kinds of movement.

Now, at last, I can say how a complementary pair of propositions can divide the square unequally. Let P be any proposition represented by a region within the square. Let $P_1, P_2, \ldots, P_n$ be a set of propositions, pairwise contraries, each of which entails P;
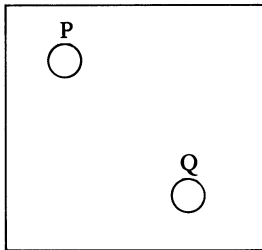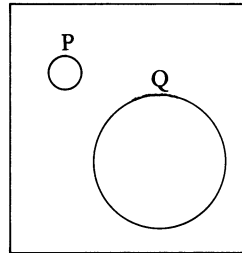

FIGURE 4   FIGURE 5

so they are propositions which state various different more specific ways in which P could be true. That can be represented as in Figure 6. Since those little propositions do not overlap with one another, and there are many of them contained within P's area, it follows that each has only a tiny fraction of the area occupied by P. Now it could be that one of them, P_k occupies an area equal to that of P's complement, i.e., equal to that part of the square which lies outside P, as in Figure 7. This would be so if P_k and $\bar{P}$ attributed different but equally specific modes of motion to the same object at the same time; for that, by my criterion, would make them equally contentful; from which it would follow that $\bar{P}$ was enormously much more contentful than — that is, was represented by a very much smaller region than — P. From this I infer, now bringing in as much as I can save of the Kant–Ayer idea, that relative to this framework P is negative and $\bar{P}$ is not.

Now I shall set the scene for applying this to the notion of instrumentality. I take someone to be instrumental in the obtaining of a state of affairs S if S does indeed obtain, and if the person's conduct makes the difference either between S's being impossible and its being on the cards, or between its being less than inevitable and its being inevitable; that is, it either hoists S's probability up from 0 or hoists it up to 1. It could do both at once, making the whole difference between probability $= 0$ and probability $= 1$ for the state of affairs S; and to keep my exposition simple I shall confine myself to that strong kind of instru-
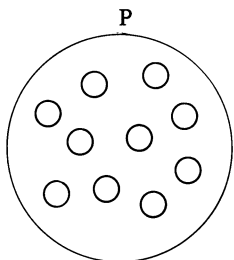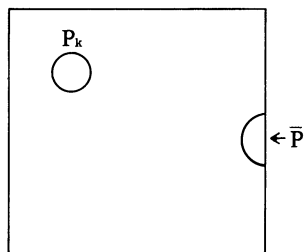

FIGURE 6   FIGURE 7

mentality. So I take it that in situation B, the vehicle is certainly all right if John doesn't dislodge the rock and certainly doomed if he does; and analogously with situation C.

If someone is in this sense instrumental in S's obtaining, his instrumentality can be simply represented on a logical-space square. W e construct a square representing all the ways the person could have moved at the relevant time, with each point in it representing one completely specific, absolutely detailed proposition. There is a unique line across the square which has on one side of it all and only the propositions which satisfy the condition:

> If it were the case that ... , S would obtain,

and on the other side of it all and only the ones satisfying the condition:

> If it were the case that ... , S would not obtain.

For example, in situations B and C there is a line through John's possible-conduct square with all and only the vehicle-is-destroyed movements on one side and all and only the vehicle-survives ones on the other.

Now, I say that in situation B — wherethe vehicle is destroyed because John dislodges the rock — what makes John's instrumentality in the vehicle's destruction positive is the fact that the line between his vehicle-is-destroyed options and his vehicle-survives ones looks like that in Figure 8. That is, of all the ways
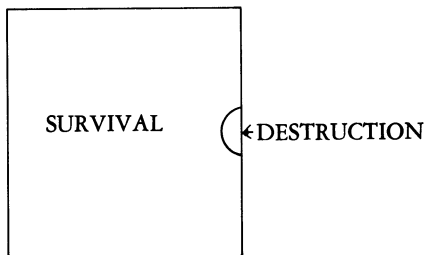

FIGURE 8

in which he could have moved, only a tiny proportion were such
as to lead to the vehicle's destruction; virtually all of them would
have had its survival as a consequence. In contrast to that, situa-
tion C looks like that in Figure 9. Here, almost any move John
could have made would have had the vehicle's destruction as a
consequence; only a tiny fragment of his possible-conduct space
contains behavioural possibilities which would have resulted in the
vehicle's surviving.

If you don't find this obvious, then here is a procedure which
should convince you that I'm right about B; and then it will be a
routine matter to adapt it to C as well. Consider the proposition
John dislodges the rock, and think about the different physical
ways he could do this: a few dozen pairwise contrary propositions
would pretty well cover the possibilities, each of them identifying
one fairly specific sort of movement which would get the rock dis-
lodged. Thus, the region of the square containing the vehicle-is-
destroyed propositions can be divided up into a few dozen still
smaller regions, each associated with some specified kind of move-
ment of one or more specified parts of John's body. Now, each
of those little propositions can be paired off with what I'll call an
'echo' of it on the vehicle-survives side of the line — that is, with
a proposition which has the same amount of content as it, and is
indeed very like it except that its truth would not in the circum-
stances result in the vehicle's destruction. For instance, if on the
vehicle-is-destroyed side we have a proposition attributing to John
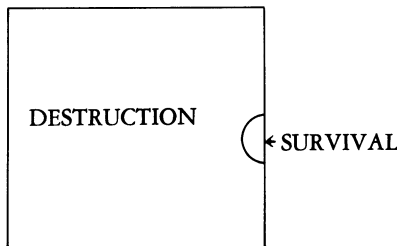

FIGURE 9

a kind of movement with his left foot, let its 'echo' be a proposition attributing to him a remarkably similar movement of that foot but with (say) the direction differing by just enough for this to be a movement which in the circumstances would not dislodge the rock and so would not doom the vehicle. In general, for each little proposition on the 'destroyed' side of the line, let its echo be one whose truth would make it look as though John were trying to dislodge the rock but had lost his sense of direction or his sense of timing or the like. Of course, the 'echo' propositions are to be pairwise contraries so that their regions don't overlap. So the drawing is like Figure 10. The pockmarks represent the echo propositions. Their combined area is the same as that of the vehicle-is-destroyed area; and it is perfectly obvious that they take up only a tiny proportion of the total vehicle-survives area. Each echo proposition attributes to John some movement which is physically rather like a rock-dislodging movement — now think of all the others which are not in the least like rock-dislodging movements!

That, then, is my case for saying that in B John is positively instrumental in the vehicle's being destroyed; and it is child's play to rerun the argument to get the conclusion that in C he is negatively instrumental in this. The facts about the use of the word 'not' (he *does* dislodge the rock, he *does not* interpose the rock) have nothing to do with it.

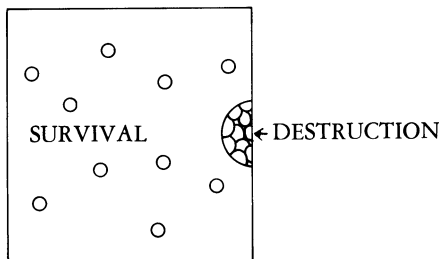This, I submit, squares pretty well with the majority of our

FIGURE 10

confident intuitions about whether someone's instrumentality is positive or negative — i.e., whether it is a case of 'because he did' or rather one of 'because he didn't'. It depends utterly on the concept of the possible *movements of a body* that provides the objective measure of degree of specificity, which is needed for two contrary propositions to have the same amount of content, which is needed in turn for one proposition to have much less content than its complement in a given square.

The restriction to bodily movements does not, so far as I know, exclude any important kind of moral situation. In assuming that the options confronting the agent are all describable in terms of how he could move, I don't think I am leaving out anything that matters.

A point which used to be much stressed is that there is no simple match-up between kinds of movement and kinds of action. But I have no need to deny that. All I need is that, given the impersonal facts of a situation, a full enough account of how the agent moves will imply an account of what he does, i.e., of what actions he performs. It just doesn't matter for what I am doing that a kind of movement which is a rock-dislodgment in one setting might be a goal kick in another.

The emphasis on bodily movement could generate the accusation that although I have established *a* distinction between positive and negative instrumentality, it turns out to be a trivial affair and not what people ordinarily have in mind when they say, for instance, that sins of omission are not as bad, other things being equal, as sins of commission. That is a serious and important charge, which will be the topic of the last part of my second lecture. I have no time to go into it now.

Another objection which has been brought against my account, and which also involves the stress on bodily movement, holds against the account something which I think is one of its merits. I shall explain.

We are to classify a person's instrumentality in S's obtaining

by looking at his possible conduct square to see how much of it is occupied by S kinds of movement and how much by non-S kinds. What about the possibility that he should make no movement at all?

In most actual situations, relative to most values of S, the person's remaining immobile would belong on the roomy side of the S/non-S line, For instance, in situation B where John dislodges the rock, if he had kept still that would have the same consequence, so far as the vehicle is concerned, as his doing any of the other things which did not get the rock dislodged. And in case C John's immobility would have been one of the many ways of not interposing the rock. In short, for most interesting values of S in most actual situations, stillness will be one way of being negatively instrumental.

But not quite always. We can construct a case where if the person stays quite still, S will inevitably ensue, whereas if he makes any movement at all non-S will certainly obtain.

Here is an example. Henry is in a sealed room where there is fine metallic dust suspended in the air. If Henry keeps utterly still for two minutes, some of the dust will settle; and if it does, some is bound to fall in such a position as to close a tiny electric circuit which . . . well, finish the story to suit your taste, but make it something big; and let's call its occurrence S. Thus, any movement from Henry, and S will not obtain; perfect immobility, and we shall get S.

I could keep my account of instrumentality silent about this case by confining it to ways of *moving* and not letting the proposition that Henry keeps still appear anywhere on his possible conduct square. Then he will not be 'instrumental' in my sense, because his ways of *moving* do not divide into non-empty S and non-S subsets, since all the ways he could move would have non-S as a consequence.

That tactic will not do. If my concept of instrumentality is to cover every sort of prima facie responsibility for upshots, then

it must cover Henry's immobility in the situation where this makes him responsible for S. So I must allow the possible-conduct squares to cover every way of moving including the null way which consists in immobility.

What my account says about Henry, then, is that if he moves he is negatively instrumental in S's obtaining, whereas if he keeps still he is positively instrumental. Henry's immobility is the sole item on the S side of our line across his possible modes of motion; everything else lies on the non-S side; and so this is a case where stillness would be a positive instrumentality and any movement would be a negative one.

I do not dispute that if Henry keeps perfectly still, he *lets* the dust fall by not creating the currents which would *keep* it in the air; but that is just one of those results which distinguishes letting from negative instrumentality and makes me want to focus on the latter.

What is there to be said for my way of drawing the line? I contend that there is everything to be said for it — i.e., that one should not think of immobility as necessarily bracketed with non-doing, non-interference, etc., and my analysis helps to show why. Why indeed should anyone think otherwise? I can think of two possible sources for this belief.

1. Someone who is infected by the idea of intrinsically negative concrete actions might be led, as Bentham was, to think that immobility is a perfect paradigm of negativeness, so that if Henry gets a result by keeping quite still we must say that this is a negative instrumentality if anything is. The idea might be that in cases where the person does move, we can call his instrumentality 'negative' in a secondary sense, meaning that, relative to the upshot we are interested in, his movements had the same result as his immobility would have had; but where he actually is immobile we have primary negativeness in all its glory. But all of this is quite worthless because it relies on the notion of negativeness *de re,* negativeness as a monadic property of concrete actions and events.

2. Some people protest that, whatever you say about instrumentality, *Henry does not move* just undeniably is a negative proposition. If there were solid theoretic grounds for saying this, I should bow out gracefully: that is, I would still distinguish instrumentalities in terms of precisely the same distinction between propositions, only I would not name the latter distinction in terms of the word 'negative'. My concern, after all, is not with that one English word, but with a whole way of looking at some material — a way which I think is profitable, however it is worded.

Still, let us ask what the credentials are for this claim that *Henry does not move* is a negative proposition.

It is clear enough, I hope, that if we are considering only the possible conduct of Henry at time T, then *Henry moves at T* is on my account a negative proposition. (I mean by that that it is negative relative to that frame of reference, i.e., is negative as compared with the complementary proposition *Henry does not move at T* My theory gives me no way of saying of any proposition that it is negative *sans phrase.)* If it seems to you insanely wrong to classify *Henry moves at T* as in any way negative, then look at it in this way. We have a square divided into tiny subregions, each representing some fairly specific way in which Henry could conduct himself with respect to motion or rest; and the proposition that Henry moves points to one tiny subregion and says *Not that one.*

What is to be said on the other side? I find no shortage of people on the other side, but a great dearth of arguments. Indeed, the only argument I have seen is one by Leibniz which so thoroughly fails as to constitute positive support for my view. It occurs in the *New Essays,* after Leibniz has quoted Locke's expressed doubts 'whether rest be any more a privation than motion', to which Leibniz responds: "I had never thought there could be any reason to doubt the privative nature of rest. All it involves is the denial of movement of the body, but it does not suffice for movement to deny rest: something else must be added to deter-

mine the degree of movement, since movement is essentially a matter of more or less whereas all states of rest are equal." But it is simply false that 'it does not suffice for movement to deny rest'. When a philosopher as good as Leibniz permits himself such a flatly false premise as that, this is evidence that he has got himself into an untenable position. On my theory of the matter, of course, he has produced the makings of an argument for classi- fying not 'x is at rest' but rather 'x is moving' as negative.

Before concluding, I want to throw a short bridge between this lecture and my next one.

The distinction which emerges from my analysis is obviously without moral significance. If someone is prima facie to blame for conduct which had a disastrous consequence, the blame could not conceivably be lessened *just* by the fact that most of his alter- native ways of behaving would have had that same consequence. That much is obvious; but of course there is more to be said, and in my second lecture I shall say some of it.

If my positive/negative distinction lacks moral significance, I can show, using that as a premise, that there is no moral sig- nificance in any of its rivals either. In conclusion, I now present that argument as applied to the distinction between doing and letting happen, with special reference to killing and letting die.

First, let us set aside positive instrumentalities which are not killings, and negative ones which are not lettings die, because the relevant probabilities are too low. If my opening the gate at the railway crossing hoists your chance of being hit by a train from 0 to 10% then if you are hit by a train I am positively instru- mental in your dying, but I have not killed you. And if my not giving you a certain medicine raises from 0 to 10% your chance of dying this week, then if you do die this week I am negatively instrumental in this, but I have not let you die, even if I have the requisite knowledge, responsibility, and so on. These matters of probability are of great moral importance, but they cannot bring moral significance to the killing/letting-die distinction

since they fall between killing and positive instrumentality in the same way as between letting die and negative instrumentality.

Next, we should set aside positive lettings die, such as letting a climber fall to his death by cutting his rope, or letting a terminal patient die by unplugging his respirator; and set aside negative killings, such as killing your baby by not feeding it. Of several reasons for setting these aside, the simplest is that in these cases the very same conduct is both a killing and a letting die; and so they cannot be in question when someone says that, other things being equal, killing is worse than letting die.

What remain to be considered are negative lettings die and positive killings. My argument on them runs as follows. My premise is that negative instrumentality in someone's dying is no better in itself than positive; and the desired conclusion is that letting someone die is no better in itself than killing him. To get from premise to conclusion I shall argue for two lemmas: that letting die is no better than the relevant negative instrumentality, and that the relevant positive instrumentality is no better than killing. Slide the premise in between those two and the conclusion rolls smoothly out.

Lettings die are negative instrumentalities marked off by special features which tend to increase moral weight and certainly do not lessen it. If I am negatively instrumental in a premature death, the addition of facts which imply that I *let* the person die will tend to make my culpability greater, not less; for they are facts such as that I had the relevant knowledge, had some responsibility in the matter, and so on.

So much for the first lemma. The second is more complicated.

Killings of the kind we now have to consider are a species of positive instrumentality in people's dying; they are marked off by two differentiae which must be examined separately.

One is the absence of an intervening agent. I can be positively instrumental in your dying by forcing or persuading someone else to kill you; but in that case I don't kill you. To get any moral

leverage out of this differentia, one would have to argue that there is some exculpatory force in the plea 'I didn't kill her; I merely hired someone to kill her'; and I don't think anyone will be game to defend that.

The other differentia is the absence of intervening coincidences. This is less well known than the first one; it came clear to me only quite recently, in conversation with David Lewis; and I'll need a moment or two to expound it. Start with a positive instrumentality which is also a killing, though not a typical one. I kick a rock which starts a landslide which crashes into a lake and sends out a wave of water which drowns you as you stand in the stream fishing. In this case I kill you; and if my kicking the rock raised the chance of your dying in that way either from 0 or up to 1 then I am positively instrumental in your death. Now alter the story a bit: the kicked rock starts a landslide only because it happens to coincide with a crash of thunder; the wave goes your way only because it happens to reach the junction at one of the rare moments when the control gates are set to the left; and the water catches you only because by chance that is the moment when you are hastily wading across the stream. The fact that the causal route from my movement to your death involves several intervening coincidences seems to imply that in this case I do not kill you; yet if the probabilities are right I am positively instrumental in your dying when you do. If I am wrong about killing, then this second differentia doesn't exist and my argument was completed a minute ago. If I am right, then killing someone involves more than being positively instrumental in his dying — it requires also that the causal chain run through a stable and durable structure rather than depending on intervening coincidental events. But it seems clear that this difference in itself makes no moral difference. Of course if the causal chain involves coincidences, I may have been unable to predict that my conduct would have your dying as a consequence; and predictability is a highly morally significant matter. But it is a likely concomitant of the event-coincidence

feature, and not of its essence. So there is nothing in this second differentia which supports the idea that killing is worse, other things being equal, than positive instrumentality in deaths.

Putting the bits together, then: letting die is no better than negative instrumentality, etc.; that is no better than positive instrumentality, etc.; that is no better than killing; and so it is a mistake to think that letting die is less grave, other things being equal, than killing is. The further exploration of that thesis will be the task of my second lecture.

II. OUR NEGLECT OF THE STARVING: IS IT AS BAD AS MURDER?

I hold the not altogether uncommon opinion that, other things being equal, it is no better to be negatively than to be positively instrumental in something bad's happening; your conduct is no worse if the calamity occurs because you did not do A than if it occurs because you did do B. What led me to this view, and is still my reason for it, is a certain account of what the difference between the two sorts of instrumentality amounts to when it is made clear and objective. I can also argue that if that difference lacks moral significance then so also do those between commission and omission, acting and refraining, causing and allowing or letting, and so on, except where these are used as vehicles for moral judgments already made.

But I am addressing myself directly to the positive/negative distinction and not to any of those others. If you think about letting people die, for instance, you will have to deal not only with negative instrumentality in deaths (as when someone is let die by not being fed), but also with positive instrumentality in people's not getting artificial life support (as when someone is let die by having his respirator disconnected). I cannot find that moral thinking is helped by bringing these very different matters together, as they are brought together by 'let'. Anyway, I am con-

cerned with only one of them –i.e.,with the line between poison-
ing and not feeding, or between disconnecting a respirator and
not connecting it.

I believe that that distinction is morally neutral. I mean by
this that it *never* makes a moral difference, the only alternative
being that it *always* makes one. What is in question is the moral
significance of a universal, and I don't see how the answer could
be that sometimes it has significance and sometimes it doesn't. To
claim moral significance for it is to imply that if something bad
happens in consequence of how I behave, the fact that it happened
because I did do B rather than because I did not do A is a reason
for increased severity of moral judgment on my conduct — per-
haps outweighed by other considerations, but still a *reason* for a
certain moral conclusion. And reasons are essentially universal :
what sometimes gives a reason for X always gives a reason for X.

You will easily think of counter-examples to that, but they will
concern *derivative* reasons, whereas my topic is the thesis that the
kind of instrumentality is a source of *basic* reasons for moral
judgments. For example, if a state of affairs involves a person's
being lonely and unhappy, that is a basic reason for judging it to
be a bad one and would always be a reason for such a judgment.
In contrast, if I were to stay up for half of tonight I would be
tired tomorrow; that is a reason for my not staying up, but only
because I am committed to a philosophical discussion tomorrow;
so it is a derivative and not a basic reason, and in other circum-
stances it might well have no force or might even go the other
way. With basic reasons — ones whose force does not depend
upon particular contingencies — I stand by my claim that what
is ever a reason is always one.

Some philosophers seem to take a position according to which
the positive/negative distinction, though it does not *always*
make a moral difference, *sometimes* does so in a manner which
cannot be called derivative — a manner which does not involve
its being the occasional vehicle of something else which does make

a moral difference. The idea seems to be that positive/negative can *combine* with another item which is not itself morally significant to form a morally significant whole — as a chemical compound can have a property not possessed by any of its elements. That chemical analogy, however, is not a substitute for clear explanation, and I have never seen a version of this idea which is clear enough to be discussable. Until one is given, I must perforce drop it and focus on the clash between those who think that positive/negative always makes a moral difference and those who think that it never does.

That clash ought to be fairly easy to adjudicate, since either position could be established by just one example — find just one case where positive/negative does, unaided, contribute a moral difference, and you have established the moral significance of the distinction; find just one case where it contributes nothing, and you have established its insignificance. That *is* the theoretical situation; but in practice it doesn't help much, because it is so hard (on the one side) to be sure that an agreed moral difference *is* being contributed by positive/negative alone, or (on the other) to be sure that there really is no moral difference between the two options which differ only as positive to negative.

I shall say a good deal about the former kind of difficulty. But first I should mention the latter. Tooley and others have argued for the moral neutrality of the difference between positive and negative instrumentality by adducing particular cases of matched pairs of villainies: one man poisons his wife out of hatred and greed, another's wife takes poison by accident and, out of hatred and greed, he withholds the antidote; and we are invited to judge them to be morally on a level, and to infer that the difference between positive and negative instrumentality has no moral significance. The trouble with this is that when the seas of wickedness are running so high, we cannot possibly trust ourselves to detect every little drop of moral difference.

On the other hand, we can't use examples where there is

nothing moral at stake, for the question is about whether the difference in kind of instrumentality affects adverse moral judgments. We could avoid being distracted or dazzled by the intensity of moral concern by comparing a case where the door closes because I do push it with one where it closes because I do not pull it; but that is useless because it has no initial moral content for the positive/negative difference to work on. Later on I shall produce a pair of examples which I think will overcome both halves of this difficulty; but it is not my chief present purpose to convince you of the moral neutrality of the difference between positive and negative instrumentality, and if that were my purpose I would not try to achieve it through examples.

From here on, I shall assume the neutrality thesis, as I call it, and proceed with two tasks: to explain why the thesis might seem wrong even if it is not; and to apply it to the question of whether our negative instrumentality in the deaths of people who die because we do not save them, although we could, makes us no better than murderers.

If the answer comes up that we are as bad as murderers, that will convince most of you that the neutrality thesis is wrong. It would worry me too; but I would not count it as decisive, for I do not offer the neutrality thesis as reconcilable with all the plain man's moral convictions. Just because I have an argument for it, I am prepared to let it modify my moral views and to urge you to let it modify yours.

Still, a sufficiently intolerable theorem refutes a theory, and the conclusion that we are no better than murderers in respect of starving Cambodians might be such a theorem. Something which would *certainly* refute the neutrality thesis — I think everyone thinks — would be its permitting a surgeon to kill one healthy person in order to get the organs needed for organ transplants to save five people who would otherwise die. Since this is a choice between positive instrumentality in one death and negative in five, and since five are worse than one, must we be judging that posi-

tive instrumentality in a death is worse than negative? Judith Thomson has shown that the answer is No. She compared that transplant case with one where a runaway trolley is hurtling down a hillside on a truck leading to a point where it will kill five people unless someone throws a switch which will divert it to a place where it will kill only one person, who would otherwise have been safe. This is also a choice between positive instrumentality in one death and negative in five: *modulo* the positive/negative distinction it is indiscernible from the transplant case; yet we judge it very differently.

The moral is that if you want to refute the neutrality thesis by producing a case where an undeniable moral difference must be attributed to a moral asymmetry in positive/negative, make sure that the difference doesn't come from something else instead. If your moral response to the trolley problem differs in direction, or even just in intensity, from your response to the transplant case, then the latter must have special features to which you are giving moral weight, and positive/negative may have nothing to do with your attitude.

The biggest morally significant difference which tends to accompany the positive/negative difference has to do with amount of cost taking this to cover everything which might make a given item of conduct unattractive to the person concerned: pain, difficulty, expensiveness, boringness.

The basic notion here is that of someone's *incurring cost*. A person is incurring a greater or lesser cost at a given time depending upon how far what he is then doing is painful or expensive or otherwise unattractive to him. Now, it could not be true that one tends to incur greater cost in positive instrumentalities than in negative; for at any moment when I am positively instrumental in the obtaining of one state of affairs I am negatively instrumental in the obtaining of others. For example, at the very moment when my name goes onto a bit of paper because I do write it, the door slams because I do not hold it, water stays in a glass because I do not spill it, and so on.

But we can also speak of a *task* as costly, meaning that it cannot be carried out without incurring cost. I here take 'tasks' to include plans, projects, schemes, obedience to orders and laws and rules, and so on. And tasks, unlike particular items of conduct, *can* be divided into positive and negative: if I am ordered to write my name on a bit of paper, my task is positive; if I am ordered not to grab the door, my task is negative. And positive tasks tend to be costlier than negative ones.

It is a familiar point: we all know that rules or orders which tell us not to do certain things are in general easier to obey than ones telling us what to do. And we can understand a little better why this is so by invoking my analysis of the difference between the two sorts of instrumentality. The crux is that the proposition that someone *does* do A puts his behaviour within a quite small part of the range of all the modes of conduct which are open to him at the relevant time, while the proposition that he *does not* do A puts his conduct outside such a relatively small area and into the large remainder. The picture is as shown in Figure 11, with the square representing the logical space of his possible modes of conduct, and — for instance — *He does hold the door* being represented by the tiny bit while *He does not hold the door* is represented by the rest. To tackle a positive task is to try to squeeze your behaviour into a designated small area, and there is some chance that you cannot do that without incurring some kind of cost — i.e., that every way of behaving which lies in that area
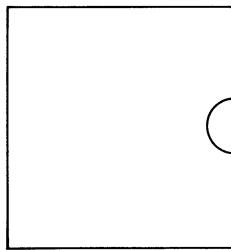

FIGURE 11

is painful or difficult or morally objectionable or whatever. On the other hand, a negative task requires you only to keep your conduct out of a certain small area and within the large remainder; and the latter's size increases the chances that you can stay within it somehow without incurring much cost. Compare being banished to Liechtenstein with being banished from Liechtenstein.

The difference in cost is quite distinct from the difference in kind of instrumentality, even if the two tend strongly to be correlated. So any views you hold about the basic morality of positive versus negative should be independent of your views about the basic morality of costs and benefits. If I can be instrumental in some good's coming to someone else, the cost to me may be too high for that amount of benefit — and perhaps too high for any amount of benefit — and this is a judgment that may be made without reference to what kind of instrumentality is involved. That is why it is wrong to say, as one writer has, that the neutrality thesis 'leads straight to an ethic so strenuous that it might give pause even to a philosophical John the Baptist.' That implies that someone who accepts the neutrality thesis must be blind to facts about cost-benefit ratios; and that is nonsense. Anyone is entitled to hold that a given cost is too high, either absolutely or for a given benefit; there is no impediment to this from the neutrality thesis.

My own practical thinking about cost-benefit ratios reflects my having an agent-relative morality: in weighing the pros and cons of a possible line of conduct, I hold myself entitled to give extra weight to a cost or benefit which is to accrue to someone with a special relation to me. The most special relation of all is identity — I am entitled to put my thumb on the moral scales on behalf of my own interests — but I extend this to my offspring, my parents, my closest friends, and so on. This is not to offer as a moral 'principle' something which refers to me in particular. What I have is a genuinely universal principle, in the manner of Kant or Hare, which allows each person to bias the scales in favour of his

own near and dear. It does not name anyone, but quantifies over everyone.

I believe that most of us have some such principle, but we probably differ a good deal in our views about which relations justify how much tilting of the moral scales, and indeed about which justify any amount of tilt — except that we will agree that identity is one of them. Without such a principle, there is no bearable answer to the question I want to put before us, namely: Does the neutrality thesis imply that we are morally no better than murderers because of our negative instrumentalities in the deaths of starving people in the third world? If we cannot use the excuse that we are only letting them die, not killing them, what excuse have we? Precious little excuse, it seems, unless we are morally entitled to tilt the moral scales on behalf of ourselves and our near and dear.

In estimating costs and benefits, I think with Bernard Williams that we should emphasise a person's 'commitments' — those long-term projects and undertakings which give his life its shape and point. In estimating the cost to someone of relinquishing some activity, for instance, we should ask not only how much good it does and how much he enjoys it, but how far it is what his life is *about,* how far it is what makes sense of him as a person with a history. This, for me as a deliberating agent, is a thought not only about *my commitments* but about *commitments:* if some plan of mine threatens not merely the comforts and pleasures of another person but the lifelong activities to which he is principally devoted, that ought to have great weight with me. My own commitments weigh more with me than his do, but only because of that general slope towards myself which also makes my pleasures count with me more than his do. Williams seems to hold that the slope is steeper with commitments than with other valuables, perhaps holding that they are not valuables at all but something deeper; but I see no reason to agree with him on either point.

The stress on commitments, incidentally, cuts across the line

between positive and negative instrumentalities. One tends to think of one's life-plans as pulling one away from positive instrumentalities, i.e., as excuses for negative ones: it is because of my devotion to philosophy that I do not take up farming, say. But they can also pull one away from negative instrumentalities, i.e., serve as excuses for positive ones: if some deeply cherished long-term project of mine requires me to act so as positively to cause harm to others, then a general stress on the importance of such projects will provide some excuse for my conduct. According to the neutrality thesis, if I can be excused for not bringing someone a benefit because to do so would jeopardise my career, then other things being equal I should also be excused for positively preventing his getting the benefit if that was needed to fend off a threat to my career. Near the end of the first chapter of *Walden,* Thoreau says: 'Probably, I should not consciously and deliberately forsake my particular calling to do the good which society demands of me, to save the universe from annihilation', thus representing himself as infinitely committed to his 'particular calling'. He won't drop it in order to spend time and energy saving the universe instead; but if his mind is clear he should also refuse to forsake his project if it is itself the universe-destroyer.

These remarks are not an argument for the neutrality thesis. They are merely a warning against thinking that by stressing the importance of deep projects and commitments you can make trouble for that thesis. The view that projects are morally crucial can be held in a form which is symmetrical with respect to the difference between positive and negative instrumentality, or in a form which is not. If you want it in the asymmetrical form you must stand up for positive/negative asymmetry as a distinct moral doctrine: you can't derive it from an emphasis on projects.

Summing up so far: if we are excusable for our relative neglect of the needy, it must be because we are entitled to give that much special weight to our own interests and those of our near and dear — with 'interest' so construed as to put great weight

on something like Williams' notion of a commitment or lifelong
project.

The conclusion I have so far come to is uncomfortable. Even
if I may tilt the moral scales in my own behalf, does this entitle-
ment go as far as the premature deaths of other people? And do
murderers have it too? Am I saying that we are as bad as mur-
derers, but that they are better than they are made out to be?
No, there is more to be said. The comparison between not giving
food to someone and shooting him dead — with which adherents
of the neutrality thesis are often triumphantly confronted — raises
too many questions at once. It cannot be usefully discussed with-
out more clearing of the undergrowth; and in the rest of this lec-
ture I shall do some of the clearing. I shall not get right back to
the comparison between ourselves and murderers, but I hope to
make that comparison easier to think about accurately.

The ensuing discussion will have another motivation which I
now explain. I shall later present a pair of cases which differ
as positive to negative and in no other way that matters, in the
hope that you will find them to be morally on a par. Really, each
is a continuum of cases, with a sliding ratio between cost to the
agent and benefit to the patient. For some cost-benefit slopes you
will condemn the conduct on both sides; for others you will
approve it on both sides; and you may agree that the break-even
point comes at about the same place on each side of the positive/
negative divide.

If my cases are to carry conviction, I must ensure that apart
from positive/negative they do not differ in any way which
might make you think — rightly or wrongly — that there is a
morally significant difference between them. I shall discuss four
such items: they have occurred to me while thinking about how
we can live with our neglect of needy people, and in presenting
them I am continuing that discussion as well as preparing for my
examples.

One concerns the fact that our moral thinking about costs and

benefits is affected by whether a cost consists in a lowering in one's level of welfare or merely in its not rising when it might have; and whether a benefit is an increase in one's welfare level or merely a non-lowering of it when it might have gone down. In contemplating a cost to me as the price of a benefit to you, it might make a difference — independently of what the amounts are — if my cost is a loss while yours is a gain, rather than, for instance, my cost being a non-gain while your benefit is a non-loss. This matter is richly present, all tangled with the difference between positive and negative instrumentality, in the language of 'harm' and 'help'. But I cannot go into that now.

This matter cannot help us to defend our neglect of the third world needy. It *might* do so if we could help only by lowering our own welfare level, and if the benefit would always be a raising of theirs, though even that would not be impressive if the levels were far enough apart to begin with. Anyway, it is not true, for many countries need first the arresting of a steady downward slide in their level of welfare. And at our end, some of us become richer as the years go by, so we could incur further costs without reducing our level of well-being, merely holding it steady when it might have risen.

The second of my four considerations concerns *rights*. If I kill a child in Cambodia, its right to life is violated, but if I let it die — do not save its life with a gift of food — I do not violate any of its rights. Or so it is usually maintained. Sometimes very bad arguments are given for this, but the conclusion is broadly correct. The concept of a moral right has a considerable positive/negative asymmetry: my right to speak is a right not to be silenced, not a right to have my laryngitis cured; my right to freedom is a right not to be imprisoned, not a right to be rescued from an underground cave; and so on. But this does not automatically refute the neutrality thesis, and so it doesn't help with our problem. If we combine the neutrality thesis with the view that the

concept of a moral right is asymmetrical as between positive and
negative, what follows is not a contradiction, but just the conclu-
sion that corresponding to any positive violation of a right there is
a possible negative instrumentality which, though it would not *be*
the violating of a right, would *be just as bad as* the violating of
a right.

That shows that in this area the concept of moral rights is not
an irresistible force. But is it perhaps an immovable object?
Suppose that someone contends that to kill a person is to violate
his right to life, whereas to let him die is neither to violate his
right to life nor to do anything as bad as that; what can I say in
reply? Really, only what I have already said in my argument for
the moral neutrality of the difference between killing and letting
die. Someone who holds that there is a morally significant differ-
ence between rights-violations and their negative analogues ought,
I contend, to show what is wrong with my argument to the con-
trary — I mean the argument given in my first lecture. If he won't
do that, that may be because he agrees that positive/negative
does not *always* make a moral difference, but holds that it can
combine with other morally inert elements to form a morally
significant compound, and that one such compound is the con-
cept of a moral right. It would. be foolish of me to say in advance
that nothing along these lines can be made to work; we can only
wait until it has been presented in enough detail for discussion of
it to be profitable.

The rights theorist might reply that he has no intention of —
and sees no need for — a general theory about how moral sig-
nificance can emerge from conceptual combinations of morally
inert elements. He would then presumably be taking his stand on
strong moral intuitions which he was prepared to retain, without
supporting theory, in the face of opposing arguments. Between
such an opponent and myself there is a fundamental disagreement
about how moral philosophy should be done — a disagreement
which I mentioned early in my first lecture and shall return to late

in my third. At this point, there is no time to do anything except salute and pass on.

That is all I want to say about rights.

In trying to make a moral comparison between positive and negative instrumentality, one runs into such facts as this. If I kill someone, e.g., by poisoning him, there is a finger pointing from me to him as the person I kill, and a finger pointing from him to me as the person who kills him. Those seem to be two large differences between this and my negative instrumentality in the deaths of people in Chad; and that pair of differences constitutes the third and fourth of my four considerations.

It seems natural to think that when someone dies in Chad, there is no finger pointing straight at me: he would have lived if I had helped him, *or* you had, *or* you *or* he *or* she had; so that my role is disjunctive in some way which reduces my responsibility.

Although that is a natural line of thought, it is just wrong. It is true that he would have lived if it had been the case that

I helped him *or* you helped him;

but now turn that right-side out, and make it say why he died, namely because

I did not help him *and* you did not help him.

The crucial statement about our negative instrumentalities is conjunctive: we were in this together, each doing his negative bit; so there is a finger pointing straight at me. There is also one pointing straight at you, but that does not lessen my responsibility. At any rate, we don't ordinarily think it does, for instance when there is a conjunction of positive instrumentalities, as when one 'provo' plants the bomb and a second detonates it. Morally speaking, there is no safety in numbers.

A simple pair of examples might help. Common to both is an electric circuit interrupted by ten switches, each controlled by a different person who has a prudential reason for wanting his

switch to be closed as much as possible. The circuit is hooked to an infernal device, so that if current flows through it an atrocity will occur. The source of the current is a dynamo over which none of the people has any control. The story then splits into two.

1. The dynamo is running, and all the switches are open. If the atrocity occurs in this case, it is because *everybody closes his switch,* i.e., because of a conjunction of positive facts about behaviour.

2. In the second version, the dynamo is about to start, and all the switches are closed. If the atrocity occurs in this case, it is because *nobody opens his switch,* i.e., because of a conjunction of negative facts about conduct.

If you are one of the ten people in case 2, you may feel that the finger of blame should not be pointed straight at you; but I contend that there are ten fingers of blame, one of which is all for you, just as in situation 1, where this is more obviously the case.

Even if you grant all this, you may wonder whether I ought to say that you are 'instrumental' in the atrocity's happening, in either of the two cases. Well, I take a person to be 'instrumental' in something's happening if his conduct put it on the cards (made the difference between probability $=0$ and probability>0) , or made it inevitable (made the difference between probability< 1 and probability $=1$). So in the case where you do not open your switch, that fact about your conduct makes the difference between there being no chance that the atrocity will occur and there being some chance. Of many changes that can be rung on this theme, I will mention only one, namely the strong kind of instrumentality in which a person's conduct makes the whole difference between probability $=0$ and probability$= 1$ for the upshot in question. The killing of a person by shooting him might be like that — an instant raising from 0 to 1 of the probability that he will die in the next few minutes — but that is not the special prerogative of positive instrumentalities. It is easy to think up negative cases which have that same feature — e.g., my second electric story in

a version where there is only one switch. But that is all by the way.

Summing up this part of the discussion, then: I hold that there is nothing morally diminishing to be said about the nature of the finger pointing towards me from a person in Chad who died when I could have saved him. Others too could have saved him, but what of that?

I now come to the last of my four considerations — the tricky matter of the finger pointing from me to the victim. For obvious reasons, I want to fix our attention on really bad things which I could have prevented. That compels me to think — as indeed I have in this talk so far — in terms of individual people whom I could have helped, especially ones whose early deaths I could have prevented.

So I do not wish to discuss my failure to give more help to needy people by contributing to programmes which would then spread my benefits as widely as possible, or my failure to institute a wide-ranging programme funded by me alone. I worked out that if I spent $1000 a month on Bangladesh, and there was no wastage or theft or administrative costs, I could, unaided, bring to every citizen of Bangladesh nearly one grain of rice a month. That's the sort of thing I don't want to discuss, because it doesn't lend itself to the sort of comparison I am trying to conduct.

So as not to lose sight of the comparison between killing and letting die, I shall focus on individual deaths which I could have prevented. In respect of each of these, my conduct made at least this much difference to whether the person died then or not: my possible ways of behaving divided into two classes, those on the 'He may die this month' side of the line and those on the 'He will live out the month' side, and I kept my conduct on the former side, giving him a probability>0 of dying in the course of the month, whereas I could have so behaved as to lower that probability pretty well to 0. And my conduct may sometimes have made the whole difference between probability $=0$ and probability $=1$.

For me to have saved all those people would have required enormous wealth, or absolute political power, or a capacity for irresistible moral persuasion. Lacking these, I could not have saved all the people. That they *all* died last year *is* a consequence of my behaviour; but that *most* of them died last year *is not,* because it was not in my power to alter that. So, one might say, each of those deaths was a disjunctive consequence of my behaviour.

It is easy for negative instrumentalities to have disjunctive consequences — states of affairs belonging to some morally homogeneous class of which one could have prevented any one but could not have prevented all. But let us be careful how the facts are expressed. In a chapter which I admire, Jonathan Glover has written: 'One difference between acts and omissions that is of some moral importance [is that] actions take time, while omissions do not. There is no end to the list of a person's omissions, while the actions he has time for during his life are limited. However heroic he is, he cannot do all the good things which, ignoring pressure of time, would be in his power.' * That is wrong. My omission to raise my arm during the past hour has taken me exactly an hour; my omission to climb Mount Everest will take me a lifetime. Also, when a number of good things are severally but not jointly in my power, the obstacle is not always one of time. It could instead be that each would require all my money, however much that might be, or that one would require me to die a virgin while the other would require me to father seven children — the obstacle to doing both of those not being a temporal one. The real reason why omissions are so numerous is that one engages in so many of them all at once: they stretch out through time just as commissions do, but they can be piled ever so much higher.

As I was saying, it often happens that we could have done

---

* Jonathan Glover, *Causing Death and Saving Lives* (New York: Penguin Books, 1977).

any one of some set of things but could not have done them all. On the positive side of the line, disjunctive consequences seem not to occur except in cases which are contrived and not interesting. I am going to assume that the disjunctive-consequence idea properly belongs only on the negative side of the line.

Since there are people of whom I could help any but not all, the question arises: 'Whom am I to help?' If it cannot be all, there must be selection. And, rightly or wrongly, we care about how the selection is made: we resist letting our benefits go arbitrarily to one person or group rather than to another. I suggest that this explains why it is sometimes said that one is more strongly obliged to help 'a starving person whom one meets on the street' than to help 'a starving person in a distant land' or — as I have heard a philosopher argue — more strongly obliged to help someone known by acquaintance than someone whom one knows only by description. Each of these opinions has strenuous implications for the moral effects of, say, a brief stopover in Bombay. What really underlies them, I suggest, is the idea that my obligation to help someone in need depends in part on why I am to consider him in particular as a recipient of my help. Granted that he is not related to me by kinship, friendship, or the like, my response to his need will depend in part upon whether it is laid before me by a natural, uncontrived course of events, or whether instead it is before me because I have — or some relief agency has — arbitrarily selected this from a large number of morally equivalent cases. In the former case, his need will press in on me much more strongly than in the latter.

Why are we like this? Perhaps it has something to do with how things are put to us by our consciences. The conscience is a considerably imaginative faculty, at least in some of us; our victims can visit us in the night, so to speak, filing past with promises to meet us at Philippi. The parade may be a long one; but it cannot be a disjunctive one, with each fist being shaken on the condition that the others are not; and so if the deaths of my

victims are all disjunctive consequences of my neglect, there is no parade; and this may help me to sleep soundly. You may find this disgraceful; or you may think — as I am inclined to — that it reflects something so deep in our natures that we had best accept it, align ourselves with it, build it into our moralities. But my immediate point is just that this *is* how we are, and that is a fact which I cannot ignore in constructing my promised pair of illustrative cases.

Here is the negative instrumentality case. There is a village in southern Africa which is about to lose its water as a result of damming operations higher up the river. The lives of the villagers will now be much worse unless they get a well: this will require capital for equipment and drilling, and a good annual sum for maintenance. All of this could be provided by 10% of my income from now until I retire; and I know that if I did make this sacrifice it really would bring durable benefit to the villagers, with few bad side effects. Furthermore, the village's needs have come to my attention in an unarbitrary manner: while engaged in ethological field work, I stayed in the area for long enough to become fully acquainted with the village's plight and the attempts that had been made to remedy it, though I formed no friendships with the villagers. I am rightly sure that if I don't provide the needed help, nobody will. Thus, if I don't give over my money, I am negatively instrumental in the village's downward slide.

The positive instrumentality story has the same village, with the same needs and dangers, but my relation to them is different. In this story I am threatened with a 10%. loss of income, though I can make up for it by pressing my claim to a trust fund which would bring me level again. If I do press my claim I shall succeed; if I don't press it, then the fund will go to provide and maintain a well for the village; and if the village does not get that money it will not get any. If I launch my lawsuit, then, I am positively instrumental in the village's downward slide.

In each story, up/down works in the same way — the cost to

me is a going down, the benefit to them would be a not going down. In each there will be few enough beneficiaries that each would benefit appreciably from the proposed conduct, and there is evidence that no incidental mischief would be done by it. In each, there is an unarbitrary, natural, uncontrived reason why I should be concerned with benefit to these people rather than to any others; that is the finger pointing from me to them.

What about the finger pointing from them to me? In the positive case there is one, because it is *my* action (not someone else's) which blocks the money from going to them, while in the negative case nothing picks out me (not someone else) as *the* one who let them down. It would be different — you might think — if nobody else *could* provide the particular help they needed. It would indeed make a difference, for in that case only one finger would point; but in the story as I have told it, there is a finger pointing at me, and I have argued that it doesn't affect *my* moral situation if there are also fingers pointing at others. So long as I am sure that nobody else *will* help the village, why should it matter whether I think that nobody else *can*? Anyway, the finger pointing at me is different from most, and perhaps all, the fingers pointing from the village towards other potential benefactors: the very same circumstances which make these villagers a special case for me make me a special case for them. Looking from the standpoint of my resources at the world of need, this village is picked out by the fact that an intimate knowledge of its needs has naturally come my way. Looking from the standpoint of the village's need at the world of potential benefactors, I am picked out by that very same fact.

The stories can be further detailed in various ways, yielding cost-benefit slopes of varying degrees of steepness. Perhaps my loss of income would change the direction of my life and my dominant activities; perhaps there would be a reduction in comfort, but nothing worse than that; or . . . and so on. And there is a similar scale on which we can adjust how bad life would be

for the villagers without the well and how good it would be with it.

Try the stories out for yourself, adjusting those parameters first in one way and then in another, though always in the same way for each. You may find that your moral response is about the same to each member of any given positive–negative pair — that a cost-benefit slope which makes the lawsuit disgraceful makes not handing over the money disgraceful, and that one which makes it permissible for me to keep my money to myself also permits me to press my claim to the trust fund.

To the extent that you respond in that way, this is evidence that your own moral thinking already embodies the neutrality thesis, even if you thought it didn't. I have constructed this pair of cases through a plodding attempt to eliminate every difference that might be thought morally significant except for that between positive and negative instrumentality. When that at last stands alone — not in a drenching downpour of moral disapproval as in Tooley's cases, and not in stories which are too skimpy for moral thought to get any grip on them — it doesn't look morally significant, does it?

But I am not finished, because there is a grave difficulty about this whole line of argument. Suppose that I replace the second story by one in which I have given my accountant full power of attorney, and I learn that through a misunderstanding he thinks it is right for him to sign away 10%, of my income to be sent to the village for its well; and I phone him up and tell him not to do that. In the comparison of that with the first member of the pair, in which I am handling my own money and I merely do not send the money to the village, I take it that no one thinks that there is the faintest moral significance in the difference between the two. But I do not produce this pair of cases in triumph: the victory comes so easily that one must suspect that the point has somehow been lost.

You might think that that pair of cases would be all right

except that the positive/negative difference is drowned, as it were, by the presence of the overpowering fact that in each case the money is *mine*. I don't want to explore that suggestion: owner-ship involves rights, and rights — I unfashionably think — are best avoided when one is doing fundamental moral philosophy. Anyway, it is easy to change the examples so that, although they still exhibit the trivial-seeming positive/negative contrast which I'm now talking about, they involve money which is not mine, though I can control who gets it. I shall not go into this in detail, and shall stay with the trivial-seeming pair of examples which is now before us.

I had better admit right away that I do regard these 'trivial' examples as establishing the neutrality thesis. I avoided them at first because I wanted to give you examples which would inspire more confidence; but I do in fact regard the trivial examples as constituting a perfectly good challenge to anyone who thinks there is moral significance in the difference between 'because he did' and 'because he didn't'. Still, I am not being truculent about this, because of course I still have a problem. I am unworried by the person who says that sometimes positive/negative does provide a basic reason for a moral discrimination while sometimes it doesn't, since that is a conceptual mistake about the nature of basic reasons. But I must take seriously the person who says that these trivial examples merely show that I have not been talking about the same positive/negative line as do those I have set myself against. Developing this idea, it could be said that in the 'trivial' pair what I do in each case is *not give* my money to the village, which is a negative fact about my conduct even if my way of *not giving* the money is to *tell* my accountant not to give it.

This brings me to the crux, which is the need for some objec-tive basis on which to draw the positive/negative line. Of course it cannot be done just according to whether the relevant sentences contain a word like 'not', for that yields no single line through the cases: compare 'He did stay home' with 'He did not go out'. My account of the difference, as I indicated earlier, rests on the

idea of a line which cuts very unequally through all the things
the person could have done; but that points to the need for
a metric for the area of logical space that the line is cutting
through — some way of counting 'kinds of things he could be
doing' which will let us say that there are far more of them on
one side of the line than on the other. My solution to that, in
terms of which I do all my thinking about these matters, is
earthily physical: I go by the different equally specific ways in
which the person could move his body. Most normal intuitions
about whether a fact about someone's conduct is positive or nega-
tive — whether it is really a 'he did' or a 'he didn't' — square
pretty well with this movement-of-body criterion. For example, if
my accountant is going to sign away my money, the ways I could
prevent him can be divided into a relatively small number of
equally specific kinds of movement with my pen or larynx; each
of those could be mapped off against an equally specific kind of
bodily movement which would not stop the money from being
signed over; and with all those mappings completed there would
still remain, unmapped, the vast majority of ways in which I could
have moved at that time. That is the force of saying that in
stopping him from signing the money I am doing one of the rela-
tively few things I could do which would have that effect; whereas
if I do not stop him, I am doing one of the vast majority of things
I could do which would have the money's being signed away as a
consequence. And it is what underlies my judgment that if I stop
him I am positively instrumental in the money's staying with me
and thus in the village's downward slide. In the other member
of that pair, where my accountant is not involved, a similar argu-
ment shows that if I do not sign away my money I am negatively
instrumental in the village's downward slide: this is because most
of the movements I could make would not get the money headed
away from me and towards them. That is why the pair do illus-
trate the difference between positive and negative instrumentality
in the village's downward slide.

Now, someone who finds that pair of cases trivial, and maintains that whether it is a matter of my not signing, or of my forbidding my accountant to sign, the dominant fact each time is the negative one that I do not so conduct myself that the money gets to the village — someone who says *that* owes us his account of how he is drawing the line between positive and negative instrumentality, or between act and omission, or between acting and refraining, or whatever. How can he do that?

His best chance seems to be through some variant on my idea of an uneven cut through the kinds of things I could do. He might say that what matters is the range of things I could do with regard to the money, and that these divide into such equal-sized kinds as investing it in bonds, spending it on pleasures of the flesh, giving it to my children, spending it on a swimming pool for myself, donating it to Oxfam, and giving it to that village in southern Africa. On this account of the matter, giving it to the village would just be so *conducting myself that it gets to the village,* with no special attention being paid to what physical movements are needed for this to happen, and thus with no regard to whether it would be a matter of my signing a document or rather of my not blocking my accountant's signing. I imagine that you find that an intuitively natural view of the matter; so do I. But what are our intuitions based on? What criteria guide our decision that 'so behaving that the money goes to the village' is to count as one kind of conduct, on a par with such other single kinds as 'so behaving that the money is invested in stocks in my name', and 'so behaving that the money buys me a swimming pool' and so on?

I have been unable to find any which do not themselves rest on prior moral judgments. The only positive/negative line I can find which is not defined in partly evaluative terms, and which rests on something deeper than sentences, is mine in terms of kinds of bodily movement. But others may be forthcoming. Or it may be counter-argued that the place of positive/negative in our moral

thinking, though important, does not require that it have deep roots in the objective world. I don't deceive myself that this work of mine is the end of anything; but I hope it may help to start something.

III. INTENDED AS A MEANS

In this lecture I shall exhibit some difficulties about a certain distinction which is thought important by many moralists — namely that between what you intend to come about as a means to your end and what you do not intend although you foresee that it will come about as a by-product of your means to your end. This has a role in most defences of the Doctrine of Double Effect, and is one source for the view that terror bombing is never permissible though tactical bombing may sometimes be — i.e., that it is never right to kill civilians as a means to demoralizing the enemy country, though it may sometimes be right to destroy a munitions factory as a means to reducing the enemy's military strength, knowing that the raid will also kill civilians. In the former case — so the story goes — the civilian deaths are intended as a means; in the latter they are not intended but merely foreseen as an inevitable by-product of the means; and that is supposed to make a moral difference, even if the probabilities are the same, the number of civilian deaths the same, and so on.

First, let us look at two kinds of causal structure:

1. movement ⟶ means (bad) ⟶ end (good)
2. movement ⟶ means (neutral) ⟶ end (good)
    ⟶ by-product (bad)

The item on the left is the movement the person makes — the 'basic action' whose upshots are in question. The other terms name particular events, and I add evaluations of them as a reminder of why these structures are supposed to be of moral interest.

In helping myself to that pair of diagrams, I am pretending to know more than I do about the identity of events; but I shall steer clear of problems about that by taking only examples which the diagrams do uncontroversially depict.

Some moralists say that a type 1 situation is worse than a type 2 one, but they are hard put to it to give reasons for this. A vague impression of reasons is sometimes conveyed by saying that in type 1 situations the bad is 'directly' produced while in type 2 ones it is not; but there is no good sense in which that is true. A type 2 case must admittedly have at least one event between the basic action and the bad event; but a type 1 case could also have an intermediate event, or a dozen of them for that matter. There is no essential difference between the two types in respect of what leads up to the bad event: the essential difference is in what flows from it; and it seems absurd to express that difference by saying that in one case but not the other the production of the bad event is 'direct'. Anyway, think for a moment about the claim that the tactical bomber in dropping live bombs onto the heads of the civilians does not 'directly' kill them!

A more usual position amongst those who morally contrast the two types of situation is not that type 1 is inherently worse than type 2 but that it is worse to intend to bring about a type 1 situation than to intend to bring about a type 2 one. That view about intentions is my chief topic in this lecture; but first I want to say two things about this use of it — that is, about the position of someone who forbids terror bombing but not tactical bombing because of an underlying judgment about the corresponding intentions.

My first remark is that this is a much odder position than is commonly recognized As a rule, if it is worse to intend to bring about X than to intend to bring about Y, that is because X is worse than Y; but here the moral difference is supposed to be introduced by the intention, rather than existing at the intention level only because of a difference at the level of events in the

world. Still, this is only an oddity, not an absurdity; and I shall
say no more about it.

My second remark is a warning against a misunderstanding.
The moral position we are confronted with has two elements. One
is a prohibitory rule which, for certain sorts of good and bad,
forbids us to produce the good by means of the bad but does not
forbid us to produce it by means which also produce the bad. The
other element is the moral judgment that it is worse to intend to
produce the bad as a means to the good than to foresee that the
bad will happen as a by-product of your means to the good. That
judgment is supposed to help justify the scope of the prohibitory
rule, but I do not take it as being, itself, such a rule. That is, I
do not see the position we are considering as including a rule
which forbids us to intend certain things.

In my opinion, it is a mistake to think of first-order morality —
morality for the guidance of deliberating agents — as making any
use of the concept of the deliberator's future intentions. The
morality I consult as a guide to my conduct does also guide my
intentions, but not by telling me what I may or may not intend.
It speaks to me of what I may or may not do, and of what are or
are not good reasons for various kinds of action; and in that way
it guides my intentions without speaking to me about them.

The concept of intention has a role in second-order morality,
i.e., in guiding judgments on people in respect of past actions.
How much I blame someone depends in part on his intentions in
acting; and if it is I who am in the dock then it is my intentions
that I must consider. But they are my past intentions, and I treat
them as external objects of judgment like anyone else's. Nothing
in this is remotely like consulting a moral rule which forbids me
to have such and such an intention.

Some moralists have been quite unclear about this. I have pre-
sented the view that a certain prohibitory rule is partly justified
by a moral fact about intentions, this being the best I can do for
the means–ends part of the Doctrine of Double Effect. But that

may be too charitable to some adherents of that doctrine. Some
of them, some of the time, write as though what were prohibited
is the having of that kind of intention, i.e., as though they gave
the concept of intention a place in first-order morality. One con-
sequence of that mistake will be mentioned later on, but the mis-
take as a whole is not something I can go into here.

My central concern, as I said, is with the thesis that it is worse
to intend to produce something bad, even if only as a means to
something good, than it is to foresee that the bad will result as a
by-product of one's means to the good. I am interested in this
only if it is maintained even when the degrees of good and bad
are the same, and the probabilities are the same. It is the thesis
that the terror bomber is in a worse frame of mind in intending to
kill ten thousand civilians as a means to lowering enemy morale
than the tactical bomber is in when he intends to destroy a factory
and confidently expects his raid to have the side effect of killing
ten thousand civilians. Some writers take examples where the
numbers of deaths, or the levels of probability, are different; but
I shall filter out such differences as those and look for the moral
significance of the difference in intention, taken on its own.

Let us see what truth there is in the statement that the terror
bomber does, while the tactical bomber does not, intend to pro-
duce something bad — specifically, to produce the deaths of
civilians. It must be a weaker sense of 'intend' than that given
by 'pursue as an end', i.e., as something sought for its own sake;
for neither of our bombers need regard civilian deaths as intrinsi-
cally desirable. But it must be stronger than 'foresee as an inevita-
ble upshot of one's conduct'; for both of our bombers foresee the
civilian deaths.

The only way I can see of driving a wedge between the two is
by invoking the view of intentions which is found in G. E. M.
Anscombe's book: this is now the dominant opinion in the rele-
vant parts of philosophy, and I am sure it is correct.* The core

* G. E. M. Anscombe, *Intention* (Oxford: Blackwell, 1957).

of it is the idea that intentions are explanatory of conduct: what
you intend is determined by which of your beliefs explain or give
your reasons for your behaviour. That immediately distinguishes
our two bombers, for the terror bomber is in some way motivated
by his expectation that his raid will produce civilian deaths, while
the tactical bomber, though having similar expectations, is in no
way motivated by them.

But let us not too rapidly draw any moral conclusions. That
there is a moral difference between the states of mind of the two
bombers is not automatically established just by the fact that one
of them intends something bad which the other does not intend.

There is moral significance in what a man intends as an end,
what he pursues for its own sake. It would be a bad man who
wanted civilian deaths for their own sakes; but neither of our
bombers is like that. This is a *sufficient* condition for intending
something, and neither bomber satisfies it.

There is also moral significance in what a man is prepared
knowingly to bring about. As Aquinas said, in effect: 'If a man
wills a bombing raid from which he knows civilian deaths will
result, it follows that he wills those deaths. Although perhaps
he does not intend the deaths in themselves, nevertheless he rather
wishes that the civilians die than that the raid be called off.' And
that is highly morally significant. But this is only a *necessary* con-
dition of intention, and it applies not just to the terror bomber
who intends the deaths but also to the tactical one who does not.
The tactical bomber would rather have civilian deaths than not
have his raid, and that is something for which he needs a pretty
good excuse. So our question is left standing: is the tactical
bomber easier to excuse than the terror one? If so, it must be for
a reason which stems from the difference in what they intend, but
it is not handed to us on a plate just by the fact that the word
'intend' fits in one case but not in the other. So we shall have to
dig for it. Let us try to be more precise about what the difference
in intention amounts to.

If intentions are determined by which of the person's beliefs motivate his action, then we should be able to get at them by asking how the behaviour would have differed if the beliefs had differed in given ways. The difference between our two men should show up in their answers to the test question:

> If you had believed that there would be no civilian deaths, would you have been less likely to go through with the raid?

Specifically, the difference should show up in the terror bomber's answering Yes and the tactical bomber's answering No. I am not saying that an intention is just a disposition to be moved by certain beliefs, merely that the difference between these two intentional states is equivalent to the difference between two dispositions to be moved by beliefs. Even that is doubtless only an approximation, but I do not think its inaccuracies matter for present purposes.

The test question is a counterfactual one, and there are different ways of interpreting it. Each man is asked: Would you have been likely to behave differently if...? If what? What is the possible state of himself which he is asked to entertain, telling us how he would have behaved if he had been in that state? We know that it is to include his thinking his raid will not lead to civilian deaths; and it had better also involve whatever follows from that by virtue of his working logic, so that it won't also include his believing, for instance, that the raid *will* cause civilian deaths. Now, how else is his supposed state to differ from the frame of mind he was actually in when he launched his raid? There are three possible interpretations.

1. His supposed state is to differ from his actual one *only* in respect of the belief that there would be no civilian deaths and its logical accompaniments — in no other way. In that case, we are leaving the terror man with his belief that his raid will lower morale, and the tactical man with his belief that his raid will destroy the factory. Each of them, then, if faced with the question

'Would you in that case have called off your raid?', will answer No. So this version of the test question does not separate them.

2 . His supposed state is to differ from his actual one in the belief that there would be no civilian deaths together with whatever follows from that by virtue of his causal beliefs. On that reading of the test question, the terror bomber will answer Yes, in that case he would have cancelled his bombing raid, for he is supposing himself to believe that there would be no civilian deaths and thus no lowering of morale — for he has the causal belief that morale can't be lowered without killing civilians. But the tactical bomber will also answer Yes, he too would have called off his raid, for he is supposing himself to believe that there would be no civilian deaths and thus no destruction of the factory — for he has the causal belief that the factory can't be destroyed without killing civilians.

Of those readings of the test question, the first supposes too little change in the antecedent state, the second too much. We need something in between, and it is not hard to see what it is.

3. The bomber's supposed state is to differ from his actual one in the belief that no civilian deaths would be caused, together with whatever follows from that, by virtue of his causal beliefs, through a causally *downstream* inference. That is, the adjustments are to concern what results, not what is causally prerequired. So the terror bomber is being supposed to think that there will be no civilian deaths and therefore no lowering of enemy morale; while the tactical bomber is being supposed to think that there will be no civilian deaths, but not to think that the factory will survive — since the factory's fate is not causally downstream from the deaths of the civilians. So the terror bomber will answer Yes, while the tactical bomber will answer No, to the test question.

That is the best I can do to clarify the difference between the two states of mind. That third reading of the test question confers reasonable clarity and undeniable truth on the statement that one man does and the other does not intend to produce civilian deaths.

But it doesn't add plausibility to the claim that this makes a moral difference. Neither bomber would call off his raid if his beliefs changed only in not including the belief that it would kill civilians. Each would call it off if they changed in that way and in every way that causally follows from it. To get them apart we had to specify what causally follows downstream and not what causally follows upstream, and I cannot see why anyone should knowingly attach moral significance to that difference as it appears here.

There is obviously great moral significance in the difference between upstream and downstream from one's own conduct. From the facts about the surgeon's behaviour it is causally inferable that there is a wounding upstream from it (he is stitching up the wound); and that is no ground for complaint against him as it would be if one could infer that there was a wounding downstream from his behaviour (because he was causing it). But that is irrelevant to our question, for in each raid the civilian deaths are downstream from the bomber's basic action.

It has been suggested that there is a difference in respect of what the two men are hoping for, or what they would in the circumstances *welcome*. The terror bomber, even if he does not want civilian deaths for themselves, still wants them — is in a frame of mind where the news of the civilian deaths would be good news — whereas the tactical bomber does not want the deaths: he merely thinks they will occur.

There is truth in that, but we must pick carefully if we are to retrieve it without bringing along falsehood as well. The terror bomber will indeed be glad when he hears that many civilians have died, because he needs their deaths for his ultimate aim. But the tactical bomber will also be glad when he hears that many civilians have died, because their deaths are evidence that something has happened which he needs for his ultimate aim. Because the raid will inevitably kill many civilians if it destroys the factory, it would be bad news for the tactical bomber if he heard that few civilians had died, for that would show that something had gone

wrong — his bombs had not exploded, or had fallen in open coun-
tryside. Something which contradicts that bad news is good news.

There is a difference between the two welcomes of the news of
civilian deaths: one man is glad because of what will flow from
the deaths, the other is glad because of what will flow from what
must have preceded them; one is downstream glad, so to speak,
while the other is upstream and then downstream glad. But there
need be no difference in how greatly glad they will be; and so,
as far as I can see, there need be no difference which creates a
moral difference.

It is true that the tactical bomber's wish for the civilian deaths
is a reluctant one: if he could, he would destroy the factory with-
out killing civilians. But the terror bomber too, if he could, would
drop his bombs in such a way as to lower morale without killing
civilians. So there is nothing in that.

It may occur to you that there is some chance of bombing the
factory without killing civilians, whereas there is none that the
terror raid will lower morale unless civilians are killed by it. This
goes with the thought that the tactical man's regret at killing civil-
ians could generate a sane, practical desire for more precise bombing
or for a wonderful coincidence in which all the civilians happen
to be out of town at the time of the raid; whereas the terror man's
regret at killing civilians could only lead to a sigh for a miracle.
That is all true, but only because of a difference in probability
which is an accident of this example; the difference between in-
tending as a means and foreseeing as a by-product is not systemati-
cally linked to a difference in probability.

Here is another reason which has been offered as making a
moral difference between the two men. Suppose for simplicity's
sake that each case involves only the death of a single civilian-
you. The tactical bomber expects his raid to kill you; but if it
doesn't, and he sees you staggering to your feet amidst the rubble
of the factory, he may rejoice. On the other hand, if the terror
bomber sees that you have survived his raid, he has reason to drop

another bomb on you, since his purpose will be defeated if you survive. This suggests a difference in how hostile they are: if the terror bomber's plans go awry, he will use his flexibility and ingenuity in ducking and weaving his way *right up to your death;* but not so the tactical bomber.

From your point of view the two cases feel different. But that difference in feeling is hard to justify unless it reflects a difference in the probability of your death; which difference exists only if there is some chance that each bomber's expectations will turn out to be wrong. But the moral doctrine I am examining is supposed to hold even when the relevant upshots are perfectly certain, so that the question doesn't even arise of the agent's using his ingenuity to deal with breakdowns in his plans.

Anyway, why should the difference in how it feels to you reflect a moral difference between the two men? Each of them is prepared to maneuver towards your death: the tactical bomber may work to overcome political resistance to his raid, evade the defences which try to keep him away from you, solve the mechanical problem with the bomb-aiming equipment, and so on, using all his skill and ingenuity and plasticity to keep on a path which has your death on it. It is true that eventually the path to your death forks away from the path to his goal, and his ingenuity goes with the latter and not the former. But he has in common with the terror bomber that he relentlessly and ingeniously pursues, *for as long as be has any reason to,* a path with your death on it. The moral difference eludes me.

It is sometimes implied that the terror bomber is *using* people as a *means* to his end whereas the tactical bomber is not. I shan't take time to sort out that tangle. As a start on it, consider whether the tactical bomber, who is supposed not to be treating people as means, is treating them as ends!

Some writers who think there is moral significance in the distinction between doing or causing on the one hand and allowing or letting on the other believe that this invests our present distinc-

tion also with moral significance. I disagree with their premise, for reasons given in my first lecture; but even if it were true, it would not do this work, as I shall now show.

If it were to do this work, the difference between what is intended and what is foreseen would have to contain or involve the difference between what you do or cause or make to happen and what you merely let or allow to happen. Some writers seem to assume that there is not merely an involvement or intertwining but a downright equivalence between these two distinctions. I have found a moral theologian clearly implying that 'the distinction between rendering someone unconscious at the risk of killing him and killing him to render him unconscious' is the same as the distinction between 'allowing to die and killing'. Another moralist slides smoothly in the reverse direction, starting with a mention of 'what we do, rather than what we allow to happen' and moving on, as though with no change of topic, to a mention of 'what we intend, and not the whole range of things which come about as a result of what we do intentionally.'

I submit that this is a mistake. Given that you do something, or actively bring it about or make it happen, it is a further question whether you intend it as a means to your end or merely foresee it as a by-product of your means; and that further question could be asked, though a bit less happily, about something which you don't do or bring about but merely allow to happen. The two distinctions cut right across one another; the belief that they are somehow aligned or intertwined seems to me to have no truth in it whatsoever.

If you are not convinced about this, consider whether you are willing to say that the tactical bomber in dropping bombs right onto people does not kill them but merely allows them to die.

*   *   *

I am a bit more than half-way through. The lecture now changes gear.

There are problems about how to apply the notion of 'intended as a means' in particular cases, and these are worth discussing even if we think that the notion has no moral significance. For one thing, they are intrinsically interesting; for another, they turn out to have a bearing on the moral significance issue, as I shall show at the end. I turn, then, to some questions about application.

Suppose that in order to save the life of a woman in labour, a surgeon performs an operation in which he crushes the head of the unborn child, thereby killing it. Must the child's death lie within the scope of what the surgeon intended, or can we say that he intended only to change the shape of the head, the death being a foreseen but unintended by-product of the procedure?

Many moralists, especially Roman Catholic ones, do condemn the crushing of the child's head; and some of them think that to do so they must say that the surgeon in crushing the child's head intends its death. By the standard of my test question, that is just wrong: it is possible and even probable that the surgeon does not intend the child to die though he knows very well that it will. In a moment or two I shall come back to that and look at some ways of trying to get around it.

First, I would like to comment briefly on the tangle which some of these moralists have got themselves into. They accept a morality which picks out some kinds of actions as absolutely prohibited, no matter what the circumstances or consequences, one prohibited kind of action being the killing of innocent human beings. Someone who holds to that ought to condemn the killing of unborn babies, no matter with what intention: if you crush the child's head you do kill it, which is to do something of an absolutely forbidden kind; and it doesn't matter whether the death was an intended means or only a foreseen by-product of your means.

On the other hand, if the prohibition of the killing of the innocent is taken literally, it condemns things which some of these moralists want to permit. Suppose for example that the removal

of a cancerous womb will certainly lead to the death of an unborn child which might become viable if the womb were left in place for two more months; and that leaving it in place for that long will remove any chance of the mother's surviving. On any tenable account of the concept of action — or of the concept of killing — the surgeon who performs this hysterectomy thereby kills the child. The causal route from his movements to the death is longer and more complex than in the head-crushing case, but what he does is a killing for all that. One common response to this is to say that it is not prohibited because it is not an intentional killing — as though the prohibition were not on killing the innocent but only on killing them where their deaths are something you intend. That puts the concept of intention into first-order morality, where I maintain it doesn't belong; and it revives the problem of how to condemn the head-crushing operation.

I am not commenting on the view that each operation might be morally permissible, or on the view that each should be categorically condemned as a killing of an innocent human being. My topic is the middle position which condemns the head-crushing but not the hysterectomy. A prohibition of intentional killing condemns neither operation; a prohibition of killing *simpliciter* condemns both. I can find no unconfused way of driving a moral wedge between them.

That is enough about that. I want now to consider how someone might try to force the death of a child into the scope of what is intended by the surgeon who crushes its head; and I shan't say any more about the confusion which may lie behind this attempt.

Philippa Foot suggests that the moralists in question might say that the child's death is 'too close' to the intended crushing of its head to fall outside the intention, the idea being that whatever is very 'close' to what you intend is itself intended. She rightly says that someone who takes this line may 'have considerable difficulty in saying where the line is to be drawn' around what is 'too close'; but that is minor compared with the difficulty

of explaining what 'too close' means in this context, and why closeness in this sense, whatever it is, should have moral significance, In the absence of any help with this, I shall spend no longer on this proposal.

The only other move I have seen relies on first shifting from whether the surgeon intends to bring about the child's death to whether he intends to kill the child. That shift may well be legitimate, and I shall not challenge it. It prepares the way for an argument which I do want to challenge, namely one which says that if the surgeon intends his crushing of the child's head he must intend his killing of the child because the crushing *is* the killing. That is Charles Fried's line of thought, as applied not to killing but to harming, when he writes: 'It is inadmissible to say that one intends to put a bullet through a man, stab him, crush him, or blow him to atoms but does not intend to harm him. All of these things just *are* harming him.'* The crucial word 'are' is not explained. Let us consider what it could mean when used in this sort of way - as it is by other writers as well. Someone who says this could be aiming to express a necessary truth about concept-inclusion: putting a bullet through someone is harming him, it might be said, in the way that suing someone is making use of the procedures of the law, and adding up numbers is doing arithmetic. But that cannot be what is meant here; for it is too obviously false that crushing and stabbing *conceptually* or *essentially* involve harming or killing. So the claim must be a contingent one about the nature of particular actions: this head-crushing (the claim must be) is a killing, this stabbing is a harming. Some theorists of action do imply such things. Anscombe and Davidson, for instance, hold that if you do X *by* doing Y then your X-ing is identical with your Y-ing; and since the surgeon kills the child *by* crushing its head, these philosophers would say that the crushing of the head is identical with the killing of the child. That is a tenable view, but it cannot be used to bring the killing of the child

---

* Charles Fried, *Right and Wrong* (Cambridge: Harvard University Press, 1978).

within the ambit of the surgeon's intention; for on this coarse-grained account of action identity, it is shiningly clear that you can intend to do X without intending to do Y even if your X-ing is your Y-ing. His pulling of the trigger was his killing of his wife; he intended to pull the trigger; he did not intend to kill his wife. Given the failure of these moves, and the apparent success of the test question, I conclude that the surgeon need not intend the death of the child whose head he crushes. Those who want to condemn the head-crushing had better find some other reason for doing so; but then I have shown that they had better do that anyway.

But we are not out of the wood. I can cheerfully accept that the surgeon does not intend the death of the child; but the principles which led to that result also produce others which make a mockery of the whole idea of what is intended as a means. I shall explain how.

First, I must bring into the open something I have mainly left implicit. Because what someone intends to bring about is limited by what he believes, the only items that can be intended are ones about which one can have beliefs. Now, there are powerful reasons — stemming mainly from work of Kripke's, I think — for saying that you cannot have a belief about a particular future event. A thought about a concrete particular must be an *effect* of that particular or of some event of which it is the subject; and if causation always runs from past to future, no thought can be an effect of an event which has not yet occurred. Therefore, there are no de re thoughts about future events. If someone loses his temper and I say 'I knew that was going to happen', I ought not to mean that I expected *that* outburst of temper, but only that I expected *an outburst of temper* pretty much like that one. And in putting the test question to one of the bombers, we are not asking how he would have behaved if he had not expected those deaths, but only how he would have behaved if he had not expected deaths.

It follows that intentions, also, are aimed at kinds of event rather than at particular events. He didn't pull the trigger intending to bring about *that* flight of the bullet, but only intending to bring about a flight of the bullet. The terror bomber could not possibly have intended to produce *those* deaths. He may have intended to bring about deaths, or deaths of civilians, or deaths of those civilians, or deaths by fire of those civilians, and so on. The intention could be aimed at a kind which is as specific as you like, but it could not be aimed at a particular event.

Now, consider the following innocent example. A political leader takes action against a trade union, intending to bring about a month-long state of disintegration in which the various locals break off from the parent body and severally fall into further disunity. This is his intended means to the end of the union's being unable to call a strike during December. He is rightly sure that if the union falls apart for that long it will never be reconstituted, but all he cares about or intends is the one-month dissolution: if he were sure that the union would recover during January and flourish for many years, that would not reduce his motivation for moving against it. Looked at intuitively, the case is a possible one, and the test question yields an acceptable description of it- one-month dissolution intended, subsequent dissolution expected but not intended.

This politician has killed the union, knowing that he was doing so. But he did not intend to kill it — if killing involves its being permanently inoperative. He intended to produce an event with a certain feature (union inoperative through December), and expected it to have a further feature (union inoperative from January first onwards); but since he didn't intend the latter, he didn't intend the conjunction of the two; and so, as I said, he did not intend to kill the union.

I am assuming that if feature F is a conjunction of features G and H, you don't intend to produce an F unless you intend to produce a G and intend to produce an H. That amounts to saying

that in delimiting what someone intends, we should shave it as close as possible. I say this because it is implied by the only account of intentions which I can make sense of. If there is a way around it — I mean one which goes deeper than an unexplained use of the phrase 'too close' — I shall listen to it with interest.

That was about the killing of a union. Now re-apply it to some killings of people — for instance the civilians in the terror raid. I said that the intention was to kill them so as to lower morale. But now that turns out to be too crude an account of the matter. All that was intended was that the people's bodies should be inoperative for long enough to cause a general belief that they were dead, this belief lasting long enough to speed the end of the war: there is nothing in that which requires, through a causally downstream inference, that the inoperativeness be permanent; and so there is nothing requiring that the people actually become dead. Of course the terror bomber knew that the people would become not merely inoperative for a while but downright dead — he had no hope of achieving the lesser thing without achieving the greater. But the greater thing is complex, and only one constituent in it was intended as a means.

There are other cases too, and not all involving death though they do all involve irreversible change: for instance, the arsonist does not intend the building to be permanently destroyed, just that it be reduced to cinders for long enough for the insurance company to pay up; and so on.

The scope of the problem can be somewhat reduced through a move suggested by Gilbert Harman. He has pointed out that the test question for delimiting intentions may make them look more fine-grained than they really are. Suppose that someone expects his behaviour to produce an F and a G, and that his only thought about this is a coarse-grained one about F-and-G-in-a-lump, with nothing in his mind corresponding to the notion of 'producing F and not producing G' — that being a mental refinement which he has not achieved. There could still be true counterfactuals about

how he would have behaved *if he had* expected to produce an F and not a G, and one of these might be that *if he had* had that expectation he would still have behaved as he did in actuality. But if his actual state of mind involves nothing of the form 'F and not G', we ought not to say that he intends to produce F and does not intend to produce G.

We could argue about what it is for a distinction to be actually registered in a mind, but let's not. It is presumably sufficient for this that the person should consciously, episodically think of the distinction — e.g., that the arsonist should play with the idea of the world's running normally to the point where the insurance company pays off, and then events in the vicinity of the building running in reverse like a film played backwards. He need not think of it as naturally possible, let alone probable. All he needs is some thought of it, sparked by fancy or whimsy or being asked the relevant test question; and similarly with the terror bomber.

So even if we give Harman's idea full force, that arsonist and that terror bomber do not intend the destruction of the building or the killing of the people. The scope of the absurdity has not been reduced much. How is the absurdity to be got rid of?

It does not help to point out that there is absolutely no chance that a disintegrated human body or incinerated building will ever be restored to full health. If that, or the terror bomber's knowledge of it, implied that he intends not only the disintegration but also its permanence, then this whole inquiry has got off on the wrong foot, as have the moralists whose views I have been exploring. They think, in my opinion rightly, that we have the notion of what is not intended but is foreseen as an *inevitable* by-product of one's means. If that is wrong, and everything which is certain to ensue from one's conduct is intended, then all we have left is the contrast between what you intend as your required means and what you foresee as a *probable* by-product of your means; and that is not something I want to discuss.

It may occur to you that when the bomber kills someone, there

is a particular, temporally circumscribed event which is the person's becoming dead: that happens at one minute past noon — it doesn't drag on forever. That is true, but it doesn't help with our problem. It would be relevant only if *that event* were what the bomber intended to produce; but that would be an intention aimed at a particular event, which is impossible. And if you shift over to the claim that he intended to produce *a* becoming-dead, you are saying precisely what I have given reason for denying: for I have argued — cogently, I hope — that the bomber intends to produce a dismantling and does not intend to produce a dying, though he is sure that the dismantling he produces will be a dying.

One might seek help from the fact that this difficulty arises only with irreversible change. Perhaps there are reasons for treating that in some special manner which protects it from my destructive argument — for example, reasons for not allowing ourselves, when delimiting intentions, to resolve a feature of the form 'F forever' into ones of the form 'F for a while' and 'F thereafter'. The trouble with that is that it would disqualify not only the absurd results with the reflective arsonist and terror bomber, but also the unabsurd treatment of the thoughtful politician's moves against the trade union.

Those are three failed attempts to neutralize the absurd results which I have reached. I can think of no others and am inclined to infer that the concept of what is 'intended as a means' cannot be given a firm, clear, theoretic grounding which implies what we think true and not what we think false regarding what people intend.

That could in turn have a bearing on the moral issue discussed in the middle part of this lecture. Someone who uses the concept of 'intended as a means' as a load-bearing part of his moral system is prima facie in trouble if the best available account of this concept leads to results which are clearly not intuitively acceptable. This is not the familiar problem about borderline cases, mentioned by Foot and dismissed as 'sophistical' by Anscombe on the grounds

that 'the fact of twilight does not mean that you cannot tell day from night.' The difficulty I have uncovered is less like twilight than like a blazing sun in a black, star-studded sky.

Let us consider what options are open to someone who thinks that there is nothing wrong with my derivation of the absurd results, and who nevertheless rests moral weight on the difference between what one intends as a means and what one foresees as a by-product.

One option would be to let the term 'intend' be guided by a rigorously close-shaving use of the test question, to accept the implication that my absurd results are true, and yet still to rest moral weight on the concept of 'intended as a means'. Cases like that of the two bombers could be coped with in either of two ways: by backing away from them, saying that there is no moral line to be drawn there; or by re-describing them, saying that the bad kind of event which one bomber intends to produce and the other doesn't is a lengthy disintegration of the bodies of civilians.

I would be charmed if any of the moralists in question did in that way embrace my absurd results. I would also be astonished. Anyway, it is an option about which I have nothing useful to say.

Another possibility is to argue that my treatment of the concept of intention can be replaced by a better one which does not yield absurd results in cases of irreversible change. I cannot discuss that possibility, of course, until the rival account is produced.

There appears to remain only one other possible response for someone who holds that the concept of 'intended as a means' has moral significance. It goes as follows.

'Your absurd results *are* absurd and should be rejected. Perhaps there is no rival analysis of intention which avoids them, but what of that? When it comes to determining what a person intends as his means to his end, we can settle this case by case, confident of mutual agreement (at least so far as the irreversible-change problem is concerned) and with no perturbation or difficulty. We agree that the good cases are perfectly good, the bad ones perfectly absurd. What more could you want?'

Of course that is enough to entitle the notion of 'intended as a means' to a place in our thought and language, but it doesn't entitle it to a place in our basic moral thinking — or anyway not in mine. This is a respect in which people differ: the difference came up early in my first lecture, and in the middle of my second, and now in conclusion I allude to it once more.

One form of dissent from it is easy to understand. It is the dissent of someone who holds that the source of moral truth is the utterances of some person (e.g., God) or some group of people (e.g., the society of which he is a member). If a person thinks that he must abide by somebody's judgment that it is wrong (say) to *kill* people, expressed with the use of the verb 'to kill', then for him that is a moral datum whose edges are as obscure to him as is the precise meaning of 'kill', as that word is used by his moral authority. And he might similarly have to take it from his authority that it is bad for civilian deaths to be 'intended as a means', regarding this as a fundamental moral truth although he can only interpret it through his informal idea of what a person can reasonably be said to intend.

It is not my purpose to mock this approach to morals, and I wish I had not once allowed myself to describe it as reducing morality to 'mere obedience'. Still, it does involve something like obedience, and that gives it its reason for admitting word-meanings into the foundations of morality. Perhaps other reasons are available to people with other views about the source of moral truth, though I don't know of any. Anyway, I invite attention to the split between those who will and those who won't conduct their basic moral thinking in terms which are apparently not controllable by clear, objective, deeply grounded conceptual principles. In these lectures I have presented some results which I have come to on one side of that split, but I have not argued for being on that side and not on the other. Rather, I have simply adopted certain standards for what makes a distinction fit to bear moral weight, and have argued that by those standards there is no

positive/negative difference, and no intended/foreseen difference which belongs in the load-bearing part of a moral structure. I am defenceless against anyone who would run the arguments in the contrapositive direction, saying that standards of fitness which lead to those conclusions should be rejected. Even you do think that, I hope you will also think that the issue has been worth bringing into the open.