II

HUMAN FREEDOM AND THE SELF

RODERICK M. CHISHOLM

'A staff moves a stone, and is moved by a hand, which is moved by a man.' Aristotle, *Physics*, 256a.

1. The metaphysical problem of human freedom might be summarized in the following way: Human beings are responsible agents; but this fact appears to conflict with a deterministic view of human action (the view that every event that is involved in an act is caused by some other event); and it also appears to conflict with an indeterministic view of human action (the view that the act, or some event that is essential to the act, is not caused at all.) To solve the problem, I believe, we must make somewhat far-reaching assumptions about the self or the agent—about the man who performs the act.

Perhaps it is needless to remark that, in all likelihood, it is impossible to say anything significant about this ancient problem that has not been said before.1

2. Let us consider some deed, or misdeed, that may be attributed to a responsible agent: one man, say, shot another. If the man was responsible for what he did, then, I would urge, what was to happen at the time of the shooting was something that was entirely up to the man himself. There was a moment at which it was true, both that he could have fired the shot and

The Lindley Lecture, 1964, pp. 3–15. © Copyright 1964 by the Department of Philosophy, University of Kansas. Reprinted by permission of the author and of the Department of Philosophy of the University of Kansas, Lawrence, Kansas, USA.

also that he could have refrained from firing it. And if this is so, then, even though he did fire it, he could have done something else instead. (He didn’t find himself firing the shot ‘against his will’, as we say.) I think we can say, more generally, then, that if a man is responsible for a certain event or a certain state of affairs (in our example, the shooting of another man), then that event or state of affairs was brought about by some act of his, and the act was something that was in his power either to perform or not to perform.

But now if the act which he did perform was an act that was also in his power not to perform, then it could not have been caused or determined by any event that was not itself within his power either to bring about or not to bring about. For example, if what we say he did was really something that was brought about by a second man, one who forced his hand upon the trigger, say, or who, by means of hypnosis, compelled him to perform the act, then since the act was caused by the second man it was nothing that was within the power of the first man to prevent. And precisely the same thing is true, I think, if instead of referring to a second man who compelled the first one, we speak instead of the desires and beliefs which the first man happens to have had. For if what we say he did was really something that was brought about by his own beliefs and desires, if these beliefs and desires in the particular situation in which he happened to have found himself caused him to do just what it was that we say he did do, then, since they caused it, he was unable to do anything other than just what it was that he did do. It makes no difference whether the cause of the deed was internal or external; if the cause was some state or event for which the man himself was not responsible, then he was not responsible for what we have been mistakenly calling his act. If a flood caused the poorly constructed dam to break, then, given the flood and the constitution of the dam, the break, we may say, had to occur and nothing could have happened in its place. And if the flood of desire caused the weak-willed man to give in, then he, too, had to do just what it was that he did do and he was no more responsible than was the dam for the results that followed. (It is true, of course, that if the man is responsible for the beliefs and desires that he happens to have, then he may also be responsible for the things they lead him to do. But the question now becomes: is he responsible for the beliefs and desires he happens to have? If he is, then there was a time when they were within his power either to acquire or not to acquire, and we are left, therefore, with our general point.)

One may object: But surely if there were such a thing as a man who is really good, then he would be responsible for things that he would do; yet, he would be unable to do anything other than just what it is that he does do, since, being good, he will always choose to do what is best. The answer, I
think, is suggested by a comment that Thomas Reid makes upon an ancient author. The author had said of Cato, ‘He was good because he could not be otherwise’, and Reid observes: ‘This saying, if understood literally and strictly, is not the praise of Cato, but of his constitution, which was no more the work of Cato than his existence’.2 If Cato was himself responsible for the good things that he did, then Cato, as Reid suggests, was such that, although he had the power to do what was not good, he exercised his power only for that which was good.

All of this, if it is true, may give a certain amount of comfort to those who are tender-minded. But we should remind them that it also conflicts with a familiar view about the nature of God—with the view that St. Thomas Aquinas expresses by saying that ‘every movement both of the will and of nature proceeds from God as the Prime Mover’.3 If the act of the sinner did proceed from God as the Prime Mover, then God was in the position of the second agent we just discussed—the man who forced the trigger finger, or the hypnotist—and the sinner, so-called, was not responsible for what he did. (This may be a bold assertion, in view of the history of western theology, but I must say that I have never encountered a single good reason for denying it.)

There is one standard objection to all of this and we should consider it briefly.

3. The objection takes the form of a stratagem—one designed to show that determinism (and divine providence) is consistent with human responsibility. The stratagem is one that was used by Jonathan Edwards and by many philosophers in the present century, most notably, G. E. Moore.4

One proceeds as follows: The expression

(a) He could have done otherwise,

it is argued, means no more nor less than

(b) If he had chosen to do otherwise, then he would have done otherwise.

(In place of ‘chosen’, one might say ‘tried’, ‘set out’, ‘decided’, ‘undertaken’, or ‘willed’.) The truth of statement (b), it is then pointed out, is consistent with determinism (and with divine providence); for even if all of the man’s actions were causally determined, the man could still be such that, if he had

2 Thomas Reid, Essays on the Active Powers of Man, essay iv, ch. 4 (Works, 600).
3 Summa Theologica, First Part of the Second Part, qu. vi (On the Voluntary and Involuntary).
chosen otherwise, then he would have done otherwise. What the murderers saw, let us suppose, along with his beliefs and desires, caused him to fire the shot; yet he was such that if, just then, he had chosen or decided not to fire the shot, then he would not have fired it. All of this is certainly possible. Similarly, we could say, of the dam, that the flood caused it to break and also that the dam was such that, if there had been no flood or any similar pressure, then the dam would have remained intact. And therefore, the argument proceeds, if (b) is consistent with determinism, and if (a) and (b) say the same thing, then (a) is also consistent with determinism; hence we can say that the agent could have done otherwise even though he was caused to do what he did do; and therefore determinism and moral responsibility are compatible.

Is the argument sound? The conclusion follows from the premises, but the catch, I think, lies in the first premiss—the one saying that statement (a) tells us no more nor less than what statement (b) tells us. For (b), it would seem, could be true while (a) is false. That is to say, our man might be such that, if he had chosen to do otherwise, then he would have done otherwise, and yet also such that he could not have done otherwise. Suppose, after all, that our murderer could not have chosen, or could not have decided, to do otherwise. Then the fact that he happens also to be a man such that, if he had chosen not to shoot he would not have shot, would make no difference. For if he could not have chosen not to shoot, then he could not have done anything other than just what it was that he did do. In a word: from our statement (b) above ('If he had chosen to do otherwise, then he would have done otherwise'), we cannot make an inference to (a) above ('He could have done otherwise') unless we can also assert:

(c) He could have chosen to do otherwise.

And therefore, if we must reject this third statement (c), then, even though we may be justified in asserting (b), we are not justified in asserting (a). If the man could not have chosen to do otherwise, then he would not have done otherwise—even if he was such that, if he had chosen to do otherwise, then he would have done otherwise.

The stratagem in question, then, seems to me not to work, and I would say, therefore, that the ascription of responsibility conflicts with a deterministic view of action.

4. Perhaps there is less need to argue that the ascription of responsibility also conflicts with an indeterministic view of action—with the view that the act, or some event that is essential to the act, is not caused at all. If the act—the firing of the shot—was not caused at all, if it was fortuitous or capricious,
happening so to speak out of the blue, then, presumably, no one—and nothing—was responsible for the act. Our conception of action, therefore, should be neither deterministic nor indeterministic. Is there any other possibility?

5. We must not say that every event involved in the act is caused by some other event; and we must not say that the act is something that is not caused at all. The possibility that remains, therefore, is this: We should say that at least one of the events that are involved in the act is caused, not by any other events, but by something else instead. And this something else can only be the agent—the man. If there is an event that is caused, not by other events, but by the man, then there are some events involved in the act that are not caused by other events. But if the event in question is caused by the man then it is caused and we are not committed to saying that there is something involved in the act that is not caused at all.

But this, of course, is a large consequence, implying something of considerable importance about the nature of the agent or the man.

6. If we consider only inanimate natural objects, we may say that causation, if it occurs, is a relation between events or states of affairs. The dam's breaking was an event that was caused by a set of other events—the dam being weak, the flood being strong, and so on. But if a man is responsible for a particular deed, then, if what I have said is true, there is some event, or set of events, that is caused, not by other events or states of affairs, but by the agent, whatever he may be.

I shall borrow a pair of medieval terms, using them, perhaps, in a way that is slightly different from that for which they were originally intended. I shall say that when one event or state of affairs (or set of events or states of affairs) causes some other event or state of affairs, then we have an instance of transeunt causation. And I shall say that when an agent, as distinguished from an event, causes an event or state of affairs, then we have an instance of immanent causation.

The nature of what is intended by the expression 'immanent causation' may be illustrated by this sentence from Aristotle's Physics: 'Thus, a staff moves a stone, and is moved by a hand, which is moved by a man.' (VII, 5, 256a, 6-8) If the man was responsible, then we have in this illustration a number of instances of causation—most of them transeunt but at least one of them immanent. What the staff did to the stone was an instance of transeunt causation, and thus we may describe it as a relation between events: 'the motion of the staff caused the motion of the stone.' And similarly for what the hand did to the staff: 'the motion of the hand caused the motion
of the staff'. And, as we know from physiology, there are still other events which caused the motion of the hand. Hence we need not introduce the agent at this particular point, as Aristotle does—we need not, though we may. We may say that the hand was moved by the man, but we may also say that the motion of the hand was caused by the motion of certain muscles; and we may say that the motion of the muscles was caused by certain events that took place within the brain. But some event, and presumably one of those that took place within the brain, was caused by the agent and not by any other events.

There are, of course, objections to this way of putting the matter; I shall consider the two that seem to me to be most important.

7. One may object, firstly: 'If the man does anything, then, as Aristotle's remark suggests, what he does is to move the hand. But he certainly does not do anything to his brain—he may not even know that he has a brain. And if he doesn't do anything to the brain, and if the motion of the hand was caused by something that happened within the brain, then there is no point in appealing to "immanent causation" as being something incompatible with "transeunt causation"—for the whole thing, after all, is a matter of causal relations among events or states of affairs.'

The answer to this objection, I think, is this: It is true that the agent does not do anything with his brain, or to his brain, in the sense in which he does something with his hand and does something to the staff. But from this it does not follow that the agent was not the immanent cause of something that happened within his brain.

We should note a useful distinction that has been proposed by Professor A. I. Melden—namely, the distinction between 'making something A happen' and 'doing A'.\(^5\) If I reach for the staff and pick it up, then one of the things that I do is just that—reach for the staff and pick it up. And if it is something that I do, then there is a very clear sense in which it may be said to be something that I know that I do. If you ask me, 'Are you doing something, or trying to do something, with the staff?', I will have no difficulty in finding an answer. But in doing something with the staff, I also make various things happen which are not in this same sense things that I do: I will make various air-particles move; I will free a number of blades of grass from the pressure that had been upon them; and I may cause a shadow to move from one place to another. If these are merely things that I make happen, as distinguished from things that I do, then I may know nothing whatever about them; I may not have the slightest idea that, in moving the

---

\(^5\) A. I. Melden, *Free Action* (London, 1961), especially ch. 3. Mr. Melden's own views, however, are quite the contrary of those that are proposed here.
staff, I am bringing about any such thing as the motion of air-particles, shadows, and blades of grass.

We may say, in answer to the first objection, therefore, that it is true that our agent does nothing to his brain or with his brain; but from this it does not follow that the agent is not the immanent cause of some event within his brain; for the brain event may be something which, like the motion of the air-particles, he made happen in picking up the staff. The only difference between the two cases is this: in each case, he made something happen when he picked up the staff; but in the one case—the motion of the air-particles or of the shadows—it was the motion of the staff that caused the event to happen; and in the other case—the event that took place in the brain—it was this event that caused the motion of the staff.

The point is, in a word, that whenever a man does something A, then (by ‘immanent causation’) he makes a certain cerebral event happen, and this cerebral event (by ‘transeunt causation’) makes A happen.

8. The second objection is more difficult and concerns the very concept of ‘immanent causation’, or causation by an agent, as this concept is to be interpreted here. The concept is subject to a difficulty which has long been associated with that of the prime mover unmoved. We have said that there must be some event A, presumably some cerebral event, which is caused not by any other event, but by the agent. Since A was not caused by any other event, then the agent himself cannot be said to have undergone any change or produced any other event (such as ‘an act of will’ or the like) which brought A about. But if, when the agent made A happen, there was no event involved other than A itself, no event which could be described as making A happen, what did the agent’s causation consist of? What, for example, is the difference between A’s just happening, and the agents’ causing A to happen? We cannot attribute the difference to any event that took place within the agent. And so far as the event A itself is concerned, there would seem to be no discernible difference. Thus Aristotle said that the activity of the prime mover is nothing in addition to the motion that it produces, and Suarez said that ‘the action is in reality nothing but the effect as it flows from the agent’.\(^6\) Must we conclude, then, that there is no more to the man’s action in causing event A than there is to the event A’s happening by itself? Here we would seem to have a distinction without a difference—in which case we have failed to find a via media between a deterministic and an indeterministic view of action.

\(^6\)Aristotle, *Physics*, bk. iii, ch. 3; Suarez, *Disputations Metaphysicae*, Disputation 18, s. 10.
The only answer, I think, can be this: that the difference between the man’s causing A, on the one hand, and the event A just happening, on the other, lies in the fact that, in the first case but not the second, the event A was caused and was caused by the man. There was a brain event A; the agent did, in fact, cause the brain event; but there was nothing that he did to cause it.

This answer may not entirely satisfy and it will be likely to provoke the following question: ‘But what are you really adding to the assertion that A happened when you utter the words “The agent caused A to happen”? As soon as we have put the question this way, we see, I think, that whatever difficulty we may have encountered is one that may be traced to the concept of causation generally—whether ‘immanent’ or ‘transeunt’. The problem, in other words, is not a problem that is peculiar to our conception of human action. It is a problem that must be faced by anyone who makes use of the concept of causation at all; and therefore, I would say, it is a problem for everyone but the complete indeterminist.

For the problem, as we put it, referring just to ‘immanent causation’, or causation by an agent, was this: ‘What is the difference between saying, of an event A, that A just happened and saying that someone caused A to happen?’ The analogous problem, which holds for ‘transeunt causation’, or causation by an event, is this: ‘What is the difference between saying, of two events A and B, that B happened and then A happened, and saying that B’s happening was the cause of A’s happening?’ And the only answer that one can give is this—that in the one case the agent was the cause of A’s happening and in the other case event B was the cause of A’s happening. The nature of transeunt causation is no more clear than is that of immanent causation.

9. But we may plausibly say—and there is a respectable philosophical tradition to which we may appeal—that the notion of immanent causation, or causation by an agent, is in fact more clear than that of transeunt causation, or causation by an event, and that it is only by understanding our own causal efficacy, as agents, that we can grasp the concept of cause at all. Hume may be said to have shown that we do not derive the concept of cause from what we perceive of external things. How, then, do we derive it? The most plausible suggestion, it seems to me, is that of Reid, once again: namely that ‘the conception of an efficient cause may very probably be derived from the experience we have had . . . of our own power to produce certain effects’.7 If we did not understand the concept of immanent causation, we would not understand that of transeunt causation.

7 Reid, Works. 524.
10. It may have been noted that I have avoided the term 'free will' in all of this. For even if there is such a faculty as 'the will', which somehow sets our acts going, the question of freedom, as John Locke said, is not the question 'whether the will be free'; it is the question 'whether a man be free'. For if there is a 'will', as a moving faculty, the question is whether the man is free to will to do these things that he does will to do—and also whether he is free not to will any of those things that he does will to do, and, again, whether he is free to will any of those things that he does not will to do. Jonathan Edwards tried to restrict himself to the question—'Is the man free to do what it is that he wills?'—but the answer to this question will not tell us whether the man is responsible for what it is that he does will to do. Using still another pair of medieval terms, we may say that the metaphysical problem of freedom does not concern the actus imperatus; it does not concern the question whether we are free to accomplish whatever it is that we will or set out to do; it concerns the actus elicitus, the question whether we are free to will or to set out to do those things that we do will or set out to do.

11. If we are responsible, and if what I have been trying to say is true, then we have a prerogative which some would attribute only to God: each of us, when we act, is a prime mover unmoved. In doing what we do, we cause certain events to happen, and nothing—or no one—causes us to cause those events to happen.

12. If we are thus prime movers unmoved and if our actions, or those for which we are responsible, are not causally determined, then they are not causally determined by our desires. And this means that the relation between what we want or what we desire, on the one hand, and what it is that we do, on the other, is not as simple as most philosophers would have it.

We may distinguish between what we might call the 'Hobbist approach' and what we might call the 'Kantian approach' to this question. The Hobbist approach is the one that is generally accepted at the present time, but the Kantian approach, I believe, is the one that is true. According to Hobbism, if we know, of some man, what his beliefs and desires happen to be and how strong they are, if we know what he feels certain of, what he desires more than anything else, and if we know the state of his body and what stimuli he is being subjected to, then we may deduce, logically, just what it is that he will do—or, more accurately, just what it is that he will try, set out, or undertake to do. Thus Professor Melden has said that 'the

8 Essay concerning Human Understanding, bk. ii, ch. 21.
connection between wanting and doing is logical'. But according to the Kantian approach to our problem, and this is the one that I would take, there is no such logical connection between wanting and doing, nor need there even be a causal connection. No set of statements about a man's desires, beliefs, and stimulus situation at any time implies any statement telling us what the man will try, set out, or undertake to do at that time. As Reid put it, though we may 'reason from men's motives to their actions and, in many cases, with great probability', we can never do so 'with absolute certainty'.

This means that, in one very strict sense of the terms, there can be no science of man. If we think of science as a matter of finding out what laws happen to hold, and if the statement of a law tells us what kinds of events are caused by what other kinds of events, then there will be human actions which we cannot explain by subsuming them under any laws. We cannot say, 'It is causally necessary that, given such and such desires and beliefs, and being subject to such and such stimuli, the agent will do so and so'. For at times the agent, if he chooses, may rise above his desires and do something else instead.

But all of this is consistent with saying that, perhaps more often than not, our desires do exist under conditions such that those conditions necessitate us to act. And we may also say, with Leibniz, that at other times our desires may 'incline without necessitating'.

13. Leibniz's phrase presents us with our final philosophical problem. What does it mean to say that a desire, or a motive, might 'incline without necessitating'? There is a temptation, certainly, to say that 'to incline' means to cause and that 'not to necessitate' means not to cause, but obviously we cannot have it both ways.

Nor will Leibniz's own solution do. In his letter to Coste, he puts the problem as follows: 'When a choice is proposed, for example to go out or not to go out, it is a question whether, with all the circumstances, internal and external, motives, perceptions, dispositions, impressions, passions, inclinations taken together, I am still in a contingent state, or whether I am necessitated to make the choice, for example, to go out; that is to say, whether this proposition true and determined in fact, In all these circumstances taken together I shall choose to go out, is contingent or necessary.' Leibniz's answer might be put as follows: in one sense of the

---

9 Melden, 166.
10 Reid, Works, 608, 612.
11 'Lettre à Mr. Coste de la Nécessité et de la Contingence' (1707) in Opera Philosophica, ed. Erdmann, 447-9.
terms 'necessary' and 'contingent', the proposition 'In all these circumstances taken together I shall choose to go out', may be said to be contingent and not necessary, and in another sense of these terms, it may be said to be necessary and not contingent. But the sense in which the proposition may be said to be contingent, according to Leibniz, is only this: there is no logical contradiction involved in denying the proposition. And the sense in which it may be said to be necessary is this: since 'nothing ever occurs without cause or determining reason', the proposition is causally necessary. 'Whenever all the circumstances taken together are such that the balance of deliberation is heavier on one side than on the other, it is certain and infallible that that is the side that is going to win out'. But if what we have been saying is true, the proposition 'In all these circumstances taken together I shall choose to go out', may be causally as well as logically contingent. Hence we must find another interpretation for Leibniz's statement that our motives and desires may incline us, or influence us, to choose without thereby necessitating us to choose.

Let us consider a public official who has some moral scruples but who also, as one says, could be had. Because of the scruples that he does have, he would never take any positive steps to receive a bribe—he would not actively solicit one. But his morality has its limits and he is also such that, if we were to confront him with a fait accompli or to let him see what is about to happen ($10,000 in cash is being deposited behind the garage), then he would succumb and be unable to resist. The general situation is a familiar one and this is one reason that people pray to be delivered from temptation. (It also justifies Kant's remark: 'And how many there are who may have led a long blameless life, who are only fortunate in having escaped so many temptations'. 12 Our relation to the misdeed that we contemplate may not be a matter simply of being able to bring it about or not to bring it about. As St. Anselm noted, there are at least four possibilities. We may illustrate them by reference to our public official and the event which is his receiving the bribe, in the following way: (i) he may be able to bring the event about himself (facere esse), in which case he would actively cause himself to receive the bribe; (ii) he may be able to refrain from bringing it about himself (non facere esse), in which case he would not himself do anything to insure that he receive the bribe; (iii) he may be able to do something to prevent the event from occurring (facere non esse), in which case he would make sure that the $10,000 was not left behind the garage; or (iv) he may be unable to do anything to prevent the event from occurring (non facere non esse), in which case, though he may not solicit the bribe, he would allow

himself to keep it.\textsuperscript{13} We have envisaged our official as a man who can resist the temptation to (i) but cannot resist the temptation to (iv): he can refrain from bringing the event about himself, but he cannot bring himself to do anything to prevent it.

Let us think of ‘inclination without necessitation’, then, in such terms as these. First we may contrast the two propositions:

(1) He can resist the temptation to do something in order to make $A$ happen;

(2) He can resist the temptation to allow $A$ to happen (i.e. to do nothing to prevent $A$ from happening).

We may suppose that the man has some desire to have $A$ happen and thus has a motive for making $A$ happen. His motive for making $A$ happen, I suggest, is one that \textit{necessitates} provided that, because of the motive, (1) is false; he cannot resist the temptation to do something in order to make $A$ happen. His motive for making $A$ happen is one that \textit{inclines} provided that, because of the motive, (2) is false; like our public official, he cannot bring himself to do anything to prevent $A$ from happening. And therefore we can say that this motive for making $A$ happen is one that \textit{inclines but does not necessitate} provided that, because of the motive, (1) is true and (2) is false; he can resist the temptation to make it happen but he cannot resist the temptation to allow it to happen.

\textsuperscript{13} Cf. D. P. Henry, ‘Saint Anselm’s \textit{De “Grammatico”}’, \textit{Philosophical Quarterly}, x (1960), 115–26. St. Anselm noted that (i) and (iii), respectively, may be thought of as forming the upper left and the upper right corners of a square of opposition, and (ii) and (iv) the lower left and the lower right.
Kant believed that good or bad luck should influence neither our moral judgement of a person and his actions, nor his moral assessment of himself.

The good will is not good because of what it effects or accomplishes or because of its adequacy to achieve some proposed end; it is good only because of its willing, i.e., it is good of itself. And, regarded for itself, it is to be esteemed incomparably higher than anything which could be brought about by it in favour of any inclination or even of the sum total of all inclinations. Even if it should happen that, by a particularly unfortunate fate or by the niggardly provision of a stepmotherly nature, this will should be wholly lacking in power to accomplish its purpose, and if even the greatest effort should not avail it to achieve anything of its end, and if there remained only the good will (not as a mere wish but as the summoning of all the means in our power), it would sparkle like a jewel in its own right, as something that had its full worth in itself. Usefulness or fruitlessness can neither diminish nor augment this worth.¹

He would presumably have said the same about a bad will: whether it accomplishes its evil purposes is morally irrelevant. And a course of action that would be condemned if it had a bad outcome cannot be vindicated if by luck it turns out well. There cannot be moral risk. This view seems to be wrong, but it arises in response to a fundamental problem about moral responsibility to which we possess no satisfactory solution.

The problem develops out of the ordinary conditions of moral judgement. Prior to reflection it is intuitively plausible that people cannot be morally assessed for what is not their fault, or for what is due to factors beyond their control. Such judgement is different from the evaluation of something as a good or bad thing, or state of affairs. The latter may be present in addition

¹ Foundations of the Metaphysics of Morals, s. 1, par. 3.
to moral judgement, but when we blame someone for his actions we are not merely saying it is bad that they happened, or bad that he exists: we are judging him, saying he is bad, which is different from his being a bad thing. This kind of judgement takes only a certain kind of object. Without being able to explain exactly why, we feel that the appropriateness of moral assessment is easily undermined by the discovery that the act or attribute, no matter how good or bad, is not under the person’s control. While other evaluations remain, this one seems to lose its footing. So a clear absence of control, produced by involuntary movement, physical force, or ignorance of the circumstances, excuses what is done from moral judgement. But what we do depends in many more ways than these on what is not under our control—that is not produced by a good or a bad will, in Kant’s phrase. And external influences on this broader range are not usually thought to excuse what is done from moral judgement, positive or negative.

Let me give a few examples, beginning with the type of case Kant has in mind. Whether we succeed or fail in what we try to do nearly always depends to some extent on factors beyond our control. This is true of murder, altruism, revolution, the sacrifice of certain interests for the sake of others—almost any morally important act. What has been done, and what is morally judged, is partly determined by external factors. However jewel-like the good will may be in its own right, there is a morally significant difference between rescuing someone from a burning building and dropping him from a twelfth-storey window while trying to rescue him. Similarly, there is a morally significant difference between reckless driving and manslaughter. But whether a reckless driver hits a pedestrian depends on the presence of the pedestrian at the point where he recklessly passes a red light. What we do is also limited by the opportunities and choices with which we are faced, and these are largely determined by factors beyond our control. Someone who was an officer in a concentration camp might have led a quiet and harmless life if the Nazis had never come to power in Germany. And someone who led a quiet and harmless life in Argentina might have become an officer in a concentration camp if he had not left Germany for business reasons in 1930.

I shall say more later about these and other examples. I introduce them here to illustrate a general point. Where a significant aspect of what someone does depends on factors beyond his control, yet we continue to treat him in that respect as an object of moral judgement, it can be called moral luck. Such luck can be good or bad. And the problem posed by this phenomenon, which led Kant to deny its possibility, is that the broad range of external influences here identified seems on close examination to undermine moral assessment as surely as does the narrower range of familiar excusing
conditions. If the condition of control is consistently applied, it threatens to erode most of the moral assessments we find it natural to make. The things for which people are morally judged are determined in more ways than we at first realize by what is beyond their control. And when the seemingly natural requirement of fault or responsibility is applied in light of these facts, it leaves few pre-reflective moral judgements intact. Ultimately, nothing or almost nothing about what a person does seems to be under his control.

Why not conclude, then, that the condition of control is false—that it is an initially plausible hypothesis refuted by clear counter-examples? One could in that case look instead for a more refined condition which picked out the kinds of lack of control that really undermine certain moral judgements, without yielding the unacceptable conclusion derived from the broader condition, that most or all ordinary moral judgements are illegitimate.

What rules out this escape is that we are dealing not with a theoretical conjecture but with a philosophical problem. The condition of control does not suggest itself merely as a generalization from certain clear cases. It seems correct in the further cases to which it is extended beyond the original set. When we undermine moral assessment by considering new ways in which control is absent, we are not just discovering what would follow given the general hypothesis, but are actually being persuaded that in itself the absence of control is relevant in these cases too. The erosion of moral judgement emerges not as the absurd consequence of an over-simple theory, but as a natural consequence of the ordinary idea of moral assessment, when it is applied in view of a more complete and precise account of the facts. It would therefore be a mistake to argue from the unacceptability of the conclusions to the need for a different account of the conditions of moral responsibility. The view that moral luck is paradoxical is not a mistake, ethical or logical, but a perception of one of the ways in which the intuitively acceptable conditions of moral judgement threaten to undermine it all.

It resembles the situation in another area of philosophy, the theory of knowledge. There too conditions which seem perfectly natural, and which grow out of the ordinary procedures for challenging and defending claims to knowledge, threaten to undermine all such claims if consistently applied. Most sceptical arguments have this quality: they do not depend on the imposition of arbitrarily stringent standards of knowledge, arrived at by misunderstanding, but appear to grow inevitably from the consistent application of ordinary standards.² There is a substantive parallel as well, for epistemological scepticism arises from consideration of the respects in

which our beliefs and their relation to reality depends on factors beyond our control. External and internal causes produce our beliefs. We may subject these processes to scrutiny in an effort to avoid error, but our conclusions at this next level also result, in part, from influences which we do not control directly. The same will be true no matter how far we carry the investigation. Our beliefs are always, ultimately, due to factors outside our control, and the impossibility of encompassing those factors without being at the mercy of others leads us to doubt whether we know anything. It looks as though, if any of our beliefs are true, it is pure biological luck rather than knowledge.

Moral luck is like this because while there are various respects in which the natural objects of moral assessment are out of our control or influenced by what is out of our control, we cannot reflect on these facts without losing our grip on the judgements.

There are roughly four ways in which the natural objects of moral assessment are disturbingly subject to luck. One is the phenomenon of constitutive luck—the kind of person you are, where this is not just a question of what you deliberately do, but of your inclinations, capacities, and temperament. Another category is luck in one’s circumstances—the kind of problems and situations one faces. The other two have to do with the causes and effects of action: luck in how one is determined by antecedent circumstances, and luck in the way one’s actions and projects turn out. All of them present a common problem. They are all opposed by the idea that one cannot be more culpable or estimable for anything than one is for that fraction of it which is under one’s control. It seems irrational to take or dispense credit or blame for matters over which a person has no control, or for their influence on results over which he has partial control. Such things may create the conditions for action, but action can be judged only to the extent that it goes beyond these conditions and does not just result from them.

Let us first consider luck, good and bad, in the way things turn out. Kant, in the above-quoted passage, has one example of this in mind, but the category covers a wide range. It includes the truck driver who accidentally runs over a child, the artist who abandons his wife and five children to devote himself to painting,\(^3\) and other cases in which the possibilities of

\(^3\) Such a case, modelled on the life of Gauguin, is discussed by Bernard Williams in ‘Moral Luck’ Proceedings of the Aristotelian Society, supplementary volume, 1976, 115–35 (to which the original version of this essay was a reply). He points out that though success or failure cannot be predicted in advance, Gauguin’s most basic retrospective feelings about the decision will be determined by the development of his talent. My disagreement with Williams is that his account fails to explain why such retrospective attitudes can be called moral. If success does not permit Gauguin to justify himself to others, but still determines his most basic feelings, that shows only that his most basic feelings need not be moral. It does not show that morality is subject to luck. If the retrospective judgement were moral, it would imply the truth of a hypothetical judgement
success and failure are even greater. The driver, if he is entirely without fault, will feel terrible about his role in the event, but will not have to reproach himself. Therefore, this example of agent-regret4 is not yet a case of moral bad luck. However, if the driver was guilty of even a minor degree of negligence—failing to have his brakes checked recently, for example—then if that negligence contributes to the death of the child, he will not merely feel terrible. He will blame himself for the death. And what makes this an example of moral luck is that he would have to blame himself only slightly for the negligence itself if no situation arose which required him to brake suddenly and violently to avoid hitting a child. Yet the negligence is the same in both cases, and the driver has no control over whether a child will run into his path.

The same is true at higher levels of negligence. If someone has had too much to drink and his car swerves on to the sidewalk, he can count himself morally lucky if there are no pedestrians in its path. If there were, he would be to blame for their deaths, and would probably be prosecuted for manslaughter. But if he hurts no one, although his recklessness is exactly the same, he is guilty of a far less serious legal offence and will certainly reproach himself and be reproached by others much less severely. To take another legal example, the penalty for attempted murder is less than that for successful murder—however similar the intentions and motives of the assailant may be in the two cases. His degree of culpability can depend, it would seem, on whether the victim happened to be wearing a bullet-proof vest, or whether a bird flew into the path of the bullet—matters beyond his control.

Finally, there are cases of decision under uncertainty—common in public and in private life. Anna Karenina goes off with Vronsky, Gauguin leaves his family, Chamberlain signs the Munich agreement, the Decembrists persuade the troops under their command to revolt against the czar, the American colonies declare their independence from Britain, you introduce two people in an attempt at match-making. It is tempting in all such cases to feel that some decision must be possible, in the light of what is known at the time, which will make reproach unsuitable no matter how things turn out. But this is not true; when someone acts in such ways he takes his life, or his moral position, into his hands, because how things turn out determines what he has done. It is possible also to assess the decision from the point of view of what could be known at the time, but this is not the end of the story. If the Decembrists had succeeded in overthrowing Nicholas I in 1825 and

made in advance, of the form 'If I leave my family and become a great painter, I will be justified by success; if I don't become a great painter, the act will be unforgivable.'

4 Williams's term ibid.
establishing a constitutional regime, they would be heroes. As it is, not only did they fail and pay for it, but they bore some responsibility for the terrible punishments meted out to the troops who had been persuaded to follow them. If the American Revolution had been a bloody failure resulting in greater repression, then Jefferson, Franklin and Washington would still have made a noble attempt, and might not even have regretted it on their way to the scaffold, but they would also have had to blame themselves for what they had helped to bring on their compatriots. (Perhaps peaceful efforts at reform would eventually have succeeded.) If Hitler had not overrun Europe and exterminated millions, but instead had died of a heart attack after occupying the Sudetenland, Chamberlain’s action at Munich would still have utterly betrayed the Czechs, but it would not be the great moral disaster that has made his name a household word.\(^5\)

In many cases of difficult choice the outcome cannot be foreseen with certainty. One kind of assessment of the choice is possible in advance, but another kind must await the outcome, because the outcome determines what has been done. The same degree of culpability or estimability in intention, motive, or concern is compatible with a wide range of judgements, positive or negative, depending on what happened beyond the point of decision. The *mens rea* which could have existed in the absence of any consequences does not exhaust the grounds of moral judgement. Actual results influence culpability or esteem in a large class of unquestionably ethical cases ranging from negligence through political choice.

That these are genuine moral judgements rather than expressions of temporary attitude is evident from the fact that one can say *in advance* how the moral verdict will depend on the results. If one negligently leaves the bath running with the baby in it, one will realize, as one bounds up the stairs toward the bathroom, that if the baby has drowned one has done something awful, whereas if it has not one has merely been careless. Someone who launches a violent revolution against an authoritarian regime knows that if he fails he will be responsible for much suffering that is in vain, but if he succeeds he will be justified by the outcome. I do not mean that *any* action can be retroactively justified by history. Certain things are so bad in themselves, or so risky, that no results can make them all right. Nevertheless, when moral judgement does depend on the outcome, it is objective and timeless and not dependent on a change of standpoint produced by success or failure. The judgement after the fact follows from an hypothetical

judgement that can be made beforehand, and it can be made as easily by someone else as by the agent.

From the point of view which makes responsibility dependent on control, all this seems absurd. How is it possible to be more or less culpable depending on whether a child gets into the path of one's car, or a bird into the path of one's bullet? Perhaps it is true that what is done depends on more than the agent's state of mind or intention. The problem then is, why is it not irrational to base moral assessment on what people do, in this broad sense? It amounts to holding them responsible for the contributions of fate as well as for their own—provided they have made some contribution to begin with. If we look at cases of negligence or attempt, the pattern seems to be that overall culpability corresponds to the product of mental or intentional fault and the seriousness of the outcome. Cases of decision under uncertainty are less easily explained in this way, for it seems that the overall judgement can even shift from positive to negative depending on the outcome. But here too it seems rational to subtract the effects of occurrences subsequent to the choice, that were merely possible at the time, and concentrate moral assessment on the actual decision in light of the probabilities. If the object of moral judgement is the person, then to hold him accountable for what he has done in the broader sense is akin to strict liability, which may have its legal uses but seems irrational as a moral position.

The result of such a line of thought is to pare down each act to its morally essential core, an inner act of pure will assessed by motive and intention. Adam Smith advocates such a position in The Theory of Moral Sentiments, but notes that it runs contrary to our actual judgements.

But how well soever we may seem to be persuaded of the truth of this equitable maxim, when we consider it after this manner, in abstract, yet when we come to particular cases, the actual consequences which happen to proceed from any action, have a very great effect upon our sentiments concerning its merit or demerit, and almost always either enhance or diminish our sense of both. Scarce, in any one instance, perhaps, will our sentiments be found, after examination, to be entirely regulated by this rule, which we all acknowledge ought entirely to regulate them.  

Joel Feinberg points out further that restricting the domain of moral responsibility to the inner world will not immunize it to luck. Factors beyond the agent's control, like a coughing fit, can interfere with his decisions as surely as they can with the path of a bullet from his gun.  Nevertheless the tendency to cut down the scope of moral assessment is

---

6 Pt. ii, s. 3, Introduction, par. 5.

pervasive, and does not limit itself to the influence of effects. It attempts to isolate the will from the other direction, so to speak, by separating out constitutive luck. Let us consider that next.

Kant was particularly insistent on the moral irrelevance of qualities of temperament and personality that are not under the control of the will. Such qualities as sympathy or coldness might provide the background against which obedience to moral requirements is more or less difficult, but they could not be objects of moral assessment themselves, and might well interfere with confident assessment of its proper object—the determination of the will by the motive of duty. This rules out moral judgement of many of the virtues and vices, which are states of character that influence choice but are certainly not exhausted by dispositions to act deliberately in certain ways. A person may be greedy, envious, cowardly, cold, ungenerous, unkind, vain, or conceit, but *behave* perfectly by a monumental effort of will. To possess these vices is to be unable to help having certain feelings under certain circumstances, and to have strong spontaneous impulses to act badly. Even if one controls the impulses, one still has the vice. An envious person hates the greater success of others. He can be morally condemned as envious even if he congratulates them cordially and does nothing to denigrate or spoil their success. Conceit, likewise, need not be displayed. It is fully present in someone who cannot help dwelling with secret satisfaction on the superiority of his own achievements, talents, beauty, intelligence, or virtue. To some extent such a quality may be the product of earlier choices; to some extent it may be amenable to change by current actions. But it is largely a matter of constitutive bad fortune. Yet people are morally condemned for such qualities, and esteemed for others equally beyond control of the will: they are assessed for what they are *like*.

To Kant this seems incoherent because virtue is enjoined on everyone and therefore must in principle be possible for everyone. It may be easier for some than for others, but it must be possible to achieve it by making the right choices, against whatever temperamental background. One may want to have a generous spirit, or regret not having one, but it makes no sense to condemn oneself or anyone else for a quality which is not within the control of the will. Condemnation implies that you should not be like that, not that it is unfortunate that you are.

Nevertheless, Kant's conclusion remains intuitively unacceptable. We

8 "If nature has put little sympathy in the heart of a man, and if he, though an honest man, is by temperament cold and indifferent to the sufferings of others, perhaps because he is provided with special gifts of patience and fortitude and expects or even requires that others should have the same—and such a man would certainly not be the meanest product of nature—would not he find in himself a source from which to give himself a far higher worth than he could have got by having a good-natured temperament?" (*Foundation of the Metaphysics of Morals*, s. 1, par. 11.)
may be persuaded that these moral judgements are irrational, but they reappear involuntarily as soon as the argument is over. This is the pattern throughout the subject.

The third category to consider is luck in one’s circumstances, and I shall mention it briefly. The things we are called upon to do, the moral tests we face, are importantly determined by factors beyond our control. It may be true of someone that in a dangerous situation he would behave in a cowardly or heroic fashion, but if the situation never arises, he will never have the chance to distinguish or disgrace himself in this way, and his moral record will be different.  

A conspicuous example of this is political. Ordinary citizens of Nazi Germany had an opportunity to behave heroically by opposing the regime. They also had an opportunity to behave badly, and most of them are culpable for having failed this test. But it is a test to which the citizens of other countries were not subjected, with the result that even if they, or some of them, would have behaved as badly as the Germans in like circumstances, they simply did not and therefore are not similarly culpable. Here again one is morally at the mercy of fate, and it may seem irrational upon reflection, but our ordinary moral attitudes would be unrecognizable without it. We judge people for what they actually do or fail to do, not just for what they would have done if circumstances had been different.

This form of moral determination by the actual is also paradoxical, but we can begin to see how deep in the concept of responsibility the paradox is embedded. A person can be morally responsible only for what he does; but what he does results from a great deal that he does not do; therefore he is not morally responsible for what he is and is not responsible for. (This is not a contradiction, but it is a paradox.)


Some mute inglorious Milton here may rest,
Some Cromwell, guiltless of his country’s blood.

An unusual example of circumstantial moral luck is provided by the kind of moral dilemma with which someone can be faced through no fault of his own, but which leaves him with nothing to do which is not wrong. See ch. 5 of Mortal Questions, and Bernard Williams, ‘Ethical Consistency’, Proceedings of the Aristotelian Society, supplementary volume, 1965, reprinted in Problems of the Self (Cambridge University Press, 1973), 166–86.

10 Circumstantial luck can extend to aspects of the situation other than individual behaviour. For example, during the Vietnam War even U.S. citizens who had opposed their country’s actions vigorously from the start often felt compromised by its crimes. Here they were not even responsible; there was probably nothing they could do to stop what was happening, so the feeling of being implicated may seem unintelligible. But it is nearly impossible to view the crimes of one’s own country in the same way that one views the crimes of another country, no matter how equal one’s lack of power to stop them in the two cases. One is a citizen of one of them, and has a connection with its actions (even if only through taxes that cannot be withheld)—that one does not have with the other’s. This makes it possible to be ashamed of one’s country, and to feel a victim of moral bad luck that one was an American in the 1960s.
It should be obvious that there is a connection between these problems about responsibility and control and an even more familiar problem, that of freedom of the will. That is the last type of moral luck I want to take up, though I can do no more within the scope of this essay than indicate its connection with the other types.

If one cannot be responsible for consequences of one’s acts due to factors beyond one’s control, or for antecedents of one’s acts that are properties of temperament not subject to one’s will, or for the circumstances that pose one’s moral choices, then how can one be responsible even for the stripped-down acts of the will itself, if they are the product of antecedent circumstances outside of the will’s control?

The area of genuine agency, and therefore of legitimate moral judgement, seems to shrink under this scrutiny to an extensionless point. Everything seems to result from the combined influence of factors, antecedent and posterior to action, that are not within the agent’s control. Since he cannot be responsible for them, he cannot be responsible for their results—though it may remain possible to take up the aesthetic or other evaluative analogues of the moral attitudes that are thus displaced.

It is also possible, of course, to brazen it out and refuse to accept the results, which indeed seem unacceptable as soon as we stop thinking about the arguments. Admittedly, if certain surrounding circumstances had been different, then no unfortunate consequences would have followed from a wicked intention, and no seriously culpable act would have been performed; but since the circumstances were not different, and the agent in fact succeeded in perpetrating a particularly cruel murder, that is what he did, and that is what he is responsible for. Similarly, we may admit that if certain antecedent circumstances had been different, the agent would never have developed into the sort of person who would do such a thing; but since he did develop (as the inevitable result of those antecedent circumstances) into the sort of swine he is, and into the person who committed such a murder, that is what he is blameable for. In both cases one is responsible for what one actually does—even if what one actually does depends in important ways on what is not within one’s control. This compatibilist account of our moral judgements would leave room for the ordinary conditions of responsibility—the absence of coercion, ignorance, or involuntary movement—as part of the determination of what someone has done—but it is understood not to exclude the influence of a great deal that he has not done.11

11 The corresponding position in epistemology would be that knowledge consists of true beliefs formed in certain ways, and that it does not require all aspects of the process to be under the knower’s control, actually or potentially. Both the correctness of these beliefs and the process by
The only thing wrong with this solution is its failure to explain how sceptical problems arise. For they arise not from the imposition of an arbitrary external requirement, but from the nature of moral judgement itself. Something in the ordinary idea of what someone does must explain how it can seem necessary to subtract from it anything that merely happens—even though the ultimate consequence of such subtraction is that nothing remains. And something in the ordinary idea of knowledge must explain why it seems to be undermined by any influences on belief not within the control of the subject—so that knowledge seems impossible without an impossible foundation in autonomous reason. But let us leave epistemology aside and concentrate on action, character, and moral assessment.

The problem arises, I believe, because the self which acts and is the object of moral judgement is threatened with dissolution by the absorption of its acts and impulses into the class of events. Moral judgement of a person is judgement not of what happens to him, but of him. It does not say merely that a certain event or state of affairs is fortunate or unfortunate or even terrible. It is not an evaluation of a state of the world, or of an individual as part of the world. We are not thinking just that it would be better if he were different, or did not exist, or had not done some of the things he has done. We are judging him, rather than his existence or characteristics. The effect of concentrating on the influence of what is not under his control is to make this responsible self seem to disappear, swallowed up by the order of mere events.

What, however, do we have in mind that a person must be to be the object of these moral attitudes? While the concept of agency is easily undermined, it is very difficult to give it a positive characterization. That is familiar from the literature on Free Will.

I believe that in a sense the problem has no solution, because something in the idea of agency is incompatible with actions being events, or people being things. But as the external determinants of what someone has done are gradually exposed, in their effect on consequences, character, and choice itself, it becomes gradually clear that actions are events and people things. Eventually nothing remains which can be ascribed to the responsible self, and we are left with nothing but a portion of the larger sequence of events, which can be deplored or celebrated, but not blamed or praised.

Though I cannot define the idea of the active self that is thus undermined, it is possible to say something about its sources. There is a close connection between our feelings about ourselves and our feelings about others. Guilt which they are arrived at would therefore be importantly subject to luck. The Nobel Prize is not awarded to people who turn out to be wrong, no matter how brilliant their reasoning.
and indignation, shame and contempt, pride and admiration are internal and external sides of the same moral attitudes. We are unable to view ourselves simply as portions of the world, and from inside we have a rough idea of the boundary between what is us and what is not, what we do and what happens to us, what is our personality and what is an accidental handicap. We apply the same essentially internal conception of the self to others. About ourselves we feel pride, shame, guilt, remorse—and agent-regret. We do not regard our actions and our characters merely as fortunate or unfortunate episodes—though they may also be that. We cannot simply take an external evaluative view of ourselves—of what we most essentially are and what we do. And this remains true even when we have seen that we are not responsible for our own existence, or our nature, or the choices we have to make, or the circumstances that give our acts the consequences they have. Those acts remain ours and we remain ourselves, despite the persuasiveness of the reasons that seem to argue us out of existence.

It is this internal view that we extend to others in moral judgement—when we judge them rather than their desirability or utility. We extend to others the refusal to limit ourselves to external evaluation, and we accord to them selves like our own. But in both cases this comes up against the brutal inclusion of humans and everything about them in a world from which they cannot be separated and of which they are nothing but contents. The external view forces itself on us at the same time that we resist it. One way this occurs is through the gradual erosion of what we do by the subtraction of what happens.\(^{12}\)

The inclusion of consequences in the conception of what we have done is an acknowledgement that we are parts of the world, but the paradoxical character of moral luck which emerges from this acknowledgement shows that we are unable to operate with such a view, for it leaves us with no one to be. The same thing is revealed in the appearance that determinism obliterates responsibility. Once we see an aspect of what we or someone else does as something that happens, we lose our grip on the idea that it has been done and that we can judge the doer and not just the happening. This explains why the absence of determinism is no more hospitable to the concept of agency than is its presence—a point that has been noticed often. Either way the act is viewed externally, as part of the course of events.

The problem of moral luck cannot be understood without an account of

the internal conception of agency and its special connection with the moral attitudes as opposed to other types of value. I do not have such an account. The degree to which the problem has a solution can be determined only by seeing whether in some degree the incompatibility between this conception and the various ways in which we do not control what we do is only apparent. I have nothing to offer on that topic either. But it is not enough to say merely that our basic moral attitudes toward ourselves and others are determined by what is actual; for they are also threatened by the sources of that actuality, and by the external view of action which forces itself on us when we see how everything we do belongs to a world that we have not created.