MORAL LUCK

B. A. O. Williams and T. Nagel

I—B. A. O. Williams

There has been a strain of philosophical thought which has identified the end of life as happiness, happiness as reflective tranquillity, and tranquillity as the product of self-sufficiency —for what is not in the domain of the self is not in its control, and so is subject to luck and the contingent enemies of tranquillity. The most extreme versions of this outlook in the Western tradition are certain doctrines of classical antiquity; though it is a notable fact about them that while the good man, the sage, was immune to the impact of incident luck, it was a matter of what may be called constitutive luck that one was a sage, or capable of becoming one: for the many and vulgar this was not (on the prevailing view) an available course.

The idea that one's whole life can in some such way be rendered immune to luck has perhaps rarely prevailed since (it did not prevail, for instance, in mainstream Christianity), but its place has been taken by the still powerfully influential idea that there is one basic form of value, moral value, which is immune to luck and—in the crucial term of the idea's most rigorous exponent—"unconditioned". Both the disposition to correct moral judgment, and the objects of such judgment, are on this view free from external contingency, for both are, in their related ways, the product of the unconditioned will. Anything which is the product of happy or unhappy contingency is no proper object of moral assessment, and no proper determinant of it either. Just as in the realm of character it is motive, not style, or powers, or endowment, that counts, so in action, it is not changes actually effected in the world, but intention. With these considerations there is supposed to disappear even that constitutive luck which the ancient sages were happy to benefit from; the capacity for moral agency is supposedly present to any rational agent whatever, to anyone for whom the question can even present itself. The successful

moral life, removed from considerations of birth, lucky upbringing, or indeed of the incomprehensible Grace of a non-Pelagian God, is presented as a career open not merely to the talents, but to a talent which all rational beings necessarily possess in the same degree. Such a conception has an ultimate form of justice at its heart, and that is its allure. Kantianism is only superficially repulsive—despite appearances, it offers an inducement, solace to a sense of the world's unfairness.

Any conception of "moral luck", on this view, is radically incoherent, and the fact that the phrase indeed sounds strange may express a fit, not unexpected, between that view and some of our implicit conceptions of morality. But the view is false. Morality itself cannot be rendered immune to luck: most basically, the dispositions of morality, however far back they are placed in the area of intention and motive, are as "conditioned" as anything else. This, the matter of what I have called "constitutive" luck, I shall leave entirely on one side. But there is a further issue. Even if moral value had been radically unconditioned by luck, it would not have been enough merely to exhibit it as one kind of value among others. Little would be affirmed unless moral values possessed some special, indeed supreme, kind of dignity or importance: the thought that there is a kind of value which is, unlike others, accessible to all rational agents, offers little encouragement if that kind of value is merely a last resort, the doss-house of the spirit. Rather, it must have a claim on one's most fundamental concerns as a rational agent, and in one's recognition of that, one is supposed to grasp, not only morality's immunity to luck, but one's own partial immunity to luck through morality.

It has notoriously not been easy for Kantianism to make clear what the recognition consists in.[1] But one consequence of it, at least, would be something very widely held: that anyone who is genuinely open to moral considerations must regard moral regret for his actions as the most basic form of regret there is, and (connectedly), in so far as he is rational, will not let his most basic regrets be determined by other than what he was fully responsible for, what lay within his voluntary control. In this way, though his life may be subject to luck, at the most basic level of his self-assessment as a rational agent, he will not be.

It is in this area of regret, justification, and the retrospective view of one's own actions, that I shall raise my questions. Some views of regret which I shall question (roughly that the most profound aspects of first-personal regret must attach to voluntary actions) are implied by this conception of morality, but may well not imply it, or indeed any specific view of morality, as opposed to certain conceptions of rationality. In so far as that is so, the discussion will have broader implications for the self's exposure to luck, though the examples centrally in question do essentially involve considerations of morality.

I shall use the notion of "luck" generously, undefinedly, but, I think, comprehensibly. (I hope it will be clear that when I say of something that it is a matter of luck, this is not meant to carry any implication that it is uncaused.) My procedure in general will be to invite reflection about how to think and feel about some rather less usual situations, in the light of an appeal to how we—many people—tend to think and feel about other more usual situations, not in terms of substantive moral opinions or "intuitions" but in terms of the experience of those kinds of situation. There is no suggestion that it is impossible for human beings to lack these feelings and experiences. In the case of the less usual there is only the claim that the thoughts and experiences I consider are possible, coherent, and intelligible, and that there is no ground for condemning them as irrational. In the case of the more usual, there are suggestions, with the outline of a reason for them, that unless we were to be merely confused or unreflective, life without these experiences would involve a much vaster reconstruction of our sentiments and our views of ourselves than may be supposed: supposed, in particular, by those philosophers who discuss these matters as though our experience of our own agency and the sense of our regrets not only could be tidied up to accord with a very simple image of rationality, but already had been.

Let us take first an outline example of the creative artist who turns away from definite and pressing human claims on him in order to live a life in which, as he supposes, he can pursue his art. Without feeling that we are limited by any historical facts, let us call him *Gauguin*. Gauguin might have been a man who was not at all interested in the claims on him, and simply preferred to live another life, and from that life,

and perhaps from that preference, his best paintings came. That sort of case, in which the claims of others simply have no hold on the agent, is not what concerns us now: though it serves to remind us of something related to the present concerns, that while we are sometimes guided by the notion that it would be the best of worlds in which morality were universally respected and all men were of a disposition to affirm it, we have in fact deep and persistent reasons to be grateful that that is not the world we have.

We are interested here in a narrower phenomenon, more intimate to moral thought itself. Let us take, rather, a Gauguin who is concerned about these claims and what is involved in their being neglected (we may suppose this to be grim), and that he nevertheless, in the face of that, opts for the other life. This other life he might perhaps not see very determinately under the category of realising his gifts as a painter: but to make consideration simpler, let us add that he does see it determinately in that light—it is as a life which will enable him really to be a painter that he opts for it. It will then be more clear what will count for him as eventual success in his project: at least some possible outcomes will be clear examples of success (which of course is not meant to be equivalent to recognition), however many others may be unclear.

Whether he will succeed cannot, in the nature of the case, be foreseen; we are not dealing here with the removal of an external obstacle to something which, once that is removed, will fairly predictably go through. Gauguin, in our story, is putting a great deal on a possibility which has not unequivocally declared itself. I want to explore and uphold the claim that it is possible that in such a situation the only thing that will justify his choice will be success itself. If he fails—and we shall come shortly to what, more precisely, failure may be—then he did the wrong thing, not just in the sense in which that platitudinously follows, but in the sense that having done the wrong thing in those circumstances he has no basis for the thought that he was justified in acting as he did; while if he succeeds, he does have a basis for that thought. This notion of justification, which I shall try to make clearer, is not one by which, if he succeeds, he will necessarily be able to justify himself *to others*. The reproaches of others he may never have

an answer to, in the sense of having a right that they accept or even listen to what he has to say; but if he fails, he will not even have anything to say.

The justification, if there is to be one, will be essentially retrospective. Gauguin could not do something which is often thought to be essential to rationality and to the notion of justification itself, which is to apply the justifying considerations at the time of the choice and in advance of knowing whether one was right (in the sense of its coming out right). How this can be in general, will form a major part of the discussion. First, however, we should consider a more limited question, whether there could be a moral justification in advance. A moral theorist, recognizing that some value attached to the success of Gauguin's project and hence possibly to his choice, might try to accommodate that choice within a framework of moral rules, by forming a subsidiary rule which could, before the outcome, justify that choice. What could that rule be? It could not be that one is morally justified in deciding to neglect other claims if one is a great creative artist: apart from basic doubts about its moral content, that saving clause begs the question which at the relevant time one is in no position to answer. On the other hand, ". . . if one is convinced that one is a great creative artist" will serve to make obstinacy and fatuous self-delusion conditions of justification; while ". . . if one is reasonably convinced that one is a great creative artist" is, if anything, worse. What is reasonable conviction supposed to be in such a case? Should Gauguin consult professors of art? The absurdity of such riders surely expresses an absurdity in the whole enterprise of trying to find a place for such cases within the rules.

If there cannot be a moral justification which is accessible in advance, then, according to the conception of morality which purges it of luck, there cannot be a moral justification at all. Whether there could in *any* sense be a moral justification of the Gauguin-type decision is not a question I shall try to resolve here. There are other issues that need discussion first, and I suspect that when they have been discussed, that will turn out to be a question of diminishing interest. But there is one point that needs to be mentioned. One consequence of finding a moral justification (a motive, perhaps for

trying to find one) might be thought to be that those who suffer from the decision would then have no justified, or at least correct, ground of reproach. There is no reason to think that we want that result. But there is also no obvious reason to think that it would be a consequence: one needs some very strong assumption about the nature of ethical consistency in order to deliver it.

Utilitarian formulations are not going to contribute any more to understanding these situations than do formulations in terms of rules. They can offer the thought "it is better (worse) that he did it", where the force of that is, approximately, "it is better (worse) that it happened", but this in itself does not help to a characterization of the agent's decision or its possible justification, and Utilitarianism has no special materials of its own to help in that. It has its own well-known problems, too, in spelling out the content of the "better"—on standard doctrine, Gauguin's decision would seem to have been a better thing, the more popular a painter he eventually became. But more interesting than that class of difficulty is the point that the Utilitarian perspective, not uniquely but clearly, will fail to attach importance to something which is actually important for these thoughts, the question of what "failure" may relevantly be. From the perspective of consequences, the goods or benefits for the sake of which Gauguin's choice was made either materialize in some degree, or do not materialize. But it matters considerably to the thoughts we are considering, in what way the project fails to come off, if it fails. If Gauguin sustains some injury on the way to Tahiti which prevents his ever painting again, that certainly means that his decision (supposing it now to be irreversible) was for nothing, and indeed there is nothing in the outcome to set against the other people's loss. But that train of events does not provoke the thought in question, that after all he was wrong and unjustified: he does not, and never will, know whether he was wrong. What would prove him wrong in his project would not just be that it failed, but that he failed.

This distinction shows that while Gauguin's justification is in some ways a matter of luck, it is not equally a matter of all kinds of luck. It matters how intrinsic the cause of failure is to

the project itself. The occurrence of an injury is, relative to these undertakings at least, luck of the most external and incident kind. Irreducibly, luck of this kind affects whether he will be justified or not, since if it strikes, he will not be justified. But it is too external for it to unjustify him, something which only his failure as a painter can do: yet still that is, at another level, luck, the luck of being able to be as he hoped he might be. It might be wondered whether that is *luck* at all, or, if so, whether it may not be luck of that constitutive kind which affects everything and which we have already left on one side. But it is more than that. It is not merely luck that he is such a man, but luck relative to the deliberations that went into his decision, that he turns out to be such a man: he might (epistemically) not have been. That is what sets the problem.

In some cases, though perhaps not in Gauguin's, success in such decisions might be thought not to be a matter of epistemic luck relative to the decision: there might be grounds for saying that the person who was prepared to take the decision, and was in fact right, actually knew that he would succeed, however subjectively uncertain he may have been. But even if this is right for some cases, it does not help with the problems of retrospective justification. For the concept of knowledge here is itself applied restrospectively, and while there is nothing necessarily wrong with that, it does not enable the agent at the time of his decision to make any distinctions he could not already make. As one might say, even if it did turn out in such a case that the agent did know, it was still luck, relative to the considerations available to him at the time and at the level at which he made his decision, that he should turn out to have known.

Some luck, in a decision of Gauguin's kind, is extrinsic to his project, some intrinsic; both are necessary for success, and hence for actual justification, but only the latter relates to unjustification. If we now broaden the range of cases slightly, we shall be able to see more clearly the notion of intrinsic luck. In Gauguin's case the nature of the project is such that two distinctions do, roughly, coincide: the distinction between luck intrinsic to the project, and luck extrinsic to it, and another distinction between what is, and what is not,

determined by him and by what he is. The intrinsic luck in Gauguin's case concentrates itself on virtually the one question of whether he is a genuinely gifted painter who can succeed in doing genuinely valuable work. Not all the conditions of the project's coming off lie in him, obviously, since others' actions and refrainings provide many necessary conditions of its coming off—and that is an important locus of extrinsic luck. But the conditions of its coming off which are relevant to unjustification, the locus of intrinsic luck, largely lie in him—which is not to say, of course, that they depend on his will, though some may. This rough coincidence of two distinctions is a feature of this case. But in others, the locus of intrinsic luck (intrinsic, that is to say, to the project) may lie partly outside the agent, and this is an important, and indeed the more typical, case.

Consider an almost equally schematized account of another example, that of Anna Karenina. Anna remains conscious in her life with Vronsky of the cost exacted from others, above all her son. She could have lived with that consciousness, we may suppose, if things had gone better; and relative to her state of understanding when she left Karenin, they could have gone better. As it turns out, the social situation and her own state of mind are such that the relationship with Vronsky has to carry too much weight, and the more obvious that becomes, the more it has to carry; and that I take that to be a truth not only about society but about her and Vronsky, a truth which, however inevitable Tolstoy ultimately makes it seem, could, relative to her earlier thoughts, have been otherwise. It is, in the present terms, a matter of intrinsic luck, and a failure in the heart of her project. But its locus is not by any means entirely in her, for it also lies in him.

It would have been an intrinsic failure, also, if Vronsky had actually committed suicide. But it would not have been that, but rather an extrinsic misfortune, if Vronsky had been accidentally killed: though her project would have been at an end, it would not have failed as it does fail. This difference illustrates precisely the thoughts we are concerned with. For if Anna had then committed suicide, her thought might essentially have been something like: "there is nothing more for me". But I take it that as things are, her thought in killing

herself is not just that, but relates inescapably also to the past and to what she has done. What she did she now finds insupportable, because she could have been justified only by the life she hoped for, and those hopes were not just negated, but refuted, by what happened.

It is these thoughts that I want to explore and to place in a structure which will make their sense plainer. The discussion is not in the first place directed to what we or others might say or think of these agents (though it has implications for that), but on what they can be expected coherently to think about themselves. A notion we shall be bound to use in describing their state of mind is *regret*, and there are certain things that need, first, to be said about this notion.

The constitutive thought of regret in general is something like "how much better if it had been otherwise", and the feeling can in principle apply to anything of which one can form some conception of how it might have been otherwise, together with consciousness of how things would then have been better. In this general sense of regret, what are regretted are states of affairs, and they can be regretted, in principle, by anyone who knows of them. But there is a particularly important species of regret, which I shall call "agent-regret", which a person can feel only towards his own past actions (or, at most, actions in which he regards himself as a participant). In this case, the supposed possible difference is that one might have acted otherwise, and the focus of the regret is on that possibility, the thought being formed in part by first-personal conceptions of how one might have acted otherwise. "Agent-regret" is not distinguished from regret in general solely or simply in virtue of its subject-matter. There can be cases of regret directed to one's own past actions which are not cases of agent-regret, because the past action is regarded purely externally, as one might regard anyone else's action. Agent-regret requires not merely a first-personal subject-matter, nor yet merely a particular kind of psychological content, but also a particular kind of expression, something which I hope will become a little clearer in what follows.

The sentiment of agent-regret is by no means restricted to *voluntary* agency. It can extend far beyond what one intentionally did to almost anything for which one was causally

responsible in virtue of something one intentionally did. Yet even at deeply accidental or non-voluntary levels of agency, sentiments of agent-regret are different from regret in general, such as might be felt by a spectator, and are acknowledged in our practice as being different. The lorry driver who, through no fault of his, runs over a child, will feel differently from any spectator, even a spectator next to him in the cab, except perhaps to the extent that the spectator takes on the thought that he might have prevented it, an agent's thought. Doubtless, and rightly, people will try, in comforting him, to move the driver from this state of feeling, move him indeed from where he is to something more like the place of a spectator; but it is important that this is seen as something that should need to be done, and indeed some doubt would be felt about a driver who too blandly or readily moved to that position. We feel sorry for the driver, but that sentiment co-exists with, indeed presupposes, that there is something special about his relation to this happening, something which cannot merely be eliminated by the consideration that it was not his fault. It may be still more so in cases where agency is fuller than in such an accident, though still involuntary through ignorance.

The differences between agent-regret and any felt by a spectator come out not just in thoughts and images that enter into the sentiment, but in differences of expression. The lorry-driver may act in some way which he hopes will constitute or at least symbolise some kind of recompense or restitution, and this will be an expression of his agent-regret. But the willingness to give compensation, even the recognition that one should give it, does not necessarily express agent-regret, and the preparedness to compensate can present itself at very different levels of significance in these connexions. We may recognize the need to pay compensation for damage we involuntarily cause, and yet this recognition be of an external kind, accompanied only by regret of a general kind, or by no regret at all. The general structure of these situations may merely be that it would be unfair for the sufferer to bear the cost if there is an alternative, and there is an alternative to be found in the agent whose intentional activities produced the damage as a side-effect. This area of compensation can be seen as part of the general regulation of boundary effects between agents' activities.

In such cases, the relevant consciousness of having done the harmful thing is basically that of its having happened as a consequence of one's acts, together with the thought that the cost of its happening can in the circumstances fairly be allocated to one's account. A test of whether that is an agent's state of mind in acknowledging that he should compensate is offered by the question whether from his point of view insurance cover would do at least as well. Imagine the premiums already paid (by someone else, we might add, if that helps to clarify the test): then if knowledge that the victim received insurance payments would settle any unease the agent feels, then it is for him an external case. It is an obvious and welcome consequence of this test that whether an agent can acceptably regard a given case externally is a function not only of his relations to it, but of what sort of case it is—besides the question of whether he should compensate rather than the insurance company, there is the question whether it is the sort of loss that can be compensated at all by insurance. If it is not, an agent conscious that he was unintentionally responsible for it might still feel that he should do something, not necessarily because he could actually compensate where insurance money could not, but because (if he is lucky) his actions might have some reparative significance other than compensation.

In other cases, again, there is no room for any appropriate action at all. Then only the desire to make reparation remains, with the painful consciousness that nothing can be done about it; some other action, perhaps less directed to the victims, may come to express this. What degree of such feeling is appropriate, and what attempts at reparative action or substitutes for it, are questions for particular cases, and that there is room in the area for irrational and self-punitive excess, no one is likely to deny. But equally it would be a kind of insanity never to experience sentiments of this kind towards anyone, and it would be an insane concept of rationality which insisted that a rational person never would. To insist on such a conception of rationality, moreover, would, apart from other kinds of absurdity, suggest a large falsehood: that we might, if we conducted ourselves clear-headedly enough, entirely detach ourselves from the unintentional aspects of our actions, relegating their costs to, so to speak, the insurance fund, and yet still retain our identity and character as agents. One's history

as an agent is a web in which anything that is the product of the will is surrounded and held up and partly formed by things that are not, in such a way that reflection can go only in one of two directions: either in the direction of saying that responsible agency is a fairly superficial concept, which has a limited use in harmonizing what happens, or else that it is not a superficial concept, but that it cannot ultimately be purified—if one attaches importance to the sense of what one is in terms of what one has done and what in the world one is responsible for, one must accept much that makes its claim on that sense solely in virtue of its being actual.[2]

The cases we are concerned with are, of course, cases of voluntary agency, but they share something with the involuntary cases just mentioned, for the "luck" of the agents relates to those elements which are essential to the outcome but lie outside their control, and what we are discussing is in this way a very drastic example of determination by the actual, the determination of the agent's judgment on his decision by what, beyond his will, actually occurs. Besides that, the discussion of agent-regret about the involuntary also helps us to get away from a dichotomy which is often relied on in these matters, expressed in such terms as *regret* and *remorse*, where "regret" is identified in effect as the regret of a spectator, while "remorse" is what we have called "agent-regret", but under the restriction that it applies only to the voluntary. The fact that we have agent-regret about the involuntary, and would not readily recognize a life without it (though we may think we might), shows already that there is something wrong with this dichotomy: such regret is neither mere spectator's regret, nor (by this definition) remorse.

There is a difference between agent-regret as we have so far discussed it, and the agents' feelings in the present cases. As we elicited it from the non-voluntary examples, agent-regret involved a wish on the agent's part that he had not done it: he deeply wishes that he had made that change which, had he known it, was in his power and which would have altered the outcome. But Gauguin or Anna Karenina, as we have represented them, wish they had acted otherwise only if they are unsuccessful. (At least, that wish attends their unsuccess under the simplifying assumption that their subsequent

thoughts and feelings are still essentially formed by the projects we have ascribed to them. This is an oversimplification, since evidently they might form new projects in the course of unsuccess itself; though Anna did not. I shall sustain the assumption in what follows.) Whatever feelings these agents had after their decision, but before the declaration of their success or failure, lacked the fully-developed wish to have acted otherwise—that wish comes only when failure is declared.

Regret necessarily involves a wish that things had been otherwise, for instance that one had not had to act as one did. But it does not necessarily involve the wish, all things taken together, that one had acted otherwise. An example of this, largely independent of the present issues, is offered by the cases of conflict between two courses of action each of which is morally required, where either course of action, even if it is judged to be for the best, leaves regrets—which are, in our present terms, agent-regrets about something voluntarily done.[3] We should not entirely assimilate agent-regret and the wish, all things taken together, to have acted otherwise. We must now look at some connexions of these to each other, and to certain ideas of justification. This will add the last element to our attempt to characterize our cases.

It will be helpful to contrast our cases with more straightforward cases of practical deliberation and the types of retrospective reflexion appropriate to them. We may take first the simplest cases of pure egoistic deliberation, where not only is the agent's attention confined to egoistic projects, but moral critics would agree that it is legitimately so confined. Here, in one sense the agent does not have to justify his deliberative processes, since there is no one he is answerable to; but it is usually supposed that there is some sense in which even such an agent's deliberative processes can be justified or unjustified —the sense, that is, in which his decision can be reasonable or unreasonable relative to his situation, whatever its actual outcome. Considerations bearing on this include at least the consistency of his thoughts, the rational assessment of probabilities, and the optimal ordering of actions in time.[4]

While the language of justification is used in this connexion, it is less clear than is usually assumed what its content is, and,

in particular, what the point is of an agent's being retrospectively concerned with the rationality of his decision, and not just with its success. How are we to understand the retrospective thought of one who comes to see a mismatch between his deliberations and the outcome? If he deliberates badly, and as a result of this his projects go wrong, it is easy to see *in that case* how his regret at the outcome appropriately attaches itself to his deliberations. But if he deliberates well, and things go wrong; particularly if, as sometimes happens, they would have gone better if he had deliberated worse; what is the consciousness that he was "justified" supposed to do for the disposition of his undoubted regret about how things actually turned out? His thought that he was justified seems to carry with it something like this: while he is sorry that things turned out as they did, and, in a sense corresponding to that, he wishes he had acted otherwise, at the same time he does not wish he had acted otherwise, for he stands by the processes of rational deliberation which led to what he did. Similarly with the converse phenomenon, where having made and too late discovered some mistake of deliberation, the agent is by luck successful, and indeed would have been less successful if he had done anything else. Here his gladness that he acted as he did (his lack of a wish to have acted otherwise) operates at a level at which it is compatible with such feelings as self-reproach or retrospective alarm at having acted as he did.

These observations are truisms, but it remains obscure what their real content is. Little is effected by talk of self-reproach or regret at all, still less of co-existent regret and contentment, unless some expression, at least, of such sentiments can be identified. Certainly it is not to be identified in this case with any disposition to compensate other persons, for none is affected. Connected with that, criticism by other persons would be on a different basis from criticism offered where they had a grievance, as in a case where an agent risks goods of which he is a trustee, through deliberative error or (interestingly) merely through the choice of a high-risk strategy to which he would be perfectly entitled if he were acting solely in his own interests. The trustee is not entitled to gamble with the infants' money even if any profits will certainly go to the infants,

and success itself will not remove, or start to remove, that objection. That sort of criticism is of course not appropriate in the purely egoistic case; and in fact there is no reason to think that criticism by others is more than a consequential consideration in the egoistic case, derived from others' recommendation of the virtues of rational prudence, which need to be explained first.

Granted that there is no issue of compensation to others in the purely egoistic case, the form of expression of regret seems necessarily to be, as Richards has said,[5] the agent's resolutions for his future deliberations. Regrets about his deliberations express themselves as resolves, at least, to think better next time; satisfaction with the deliberation, however disappointing the particular outcome, expresses itself in this, that he finds nothing to be *learned* from the case, and is sure that he will have no better chance of success (at a given level of payoff) next time by changing his procedures. If this is right, then the notions of regret or lack of regret at the past level of deliberative excellence make sense only in the context of a policy or disposition of rational deliberation applied to an on-going class of cases.

This is a modest enough conception—it is important to see how modest it is. It implies a class of cases sufficiently similar for deliberative practices to be translated from one to another of them; it does not imply that these cases are all conjointly the subject of deliberative reasoning. I may make a reasoned choice between alternatives of a certain kind today, and, having seen how it turns out, resolve to deal rather differently with the next choice of that kind; but I need not either engage in or resolve to engage in any deliberative reasoning which weighs the options of more than one such occasion together.[6]

In so far as the outcomes of different such situations affect one another, there is indeed pressure to say that rational deliberation should in principle consider them together. But, further, if one knew enough, any choice would be seen to affect all later ones; so it has seemed to some that the ideal limit of this process is something which is a far more ambitious extension of the modest notion of an ongoing disposition to rational deliberation: this is the model of rational deliberation as directed to a *life-plan*, in Rawls' sense, which treats all

times of one's life as of equal concern to one.[7] The theorists of this picture agree that as a matter of fact ignorance and other factors do usually make it rational to discount over remoteness in time, but these are subsequent considerations brought to a model which is that of one's life as a rectangle, so to speak, presented all at once and to be optimally filled in. This model is presented not only as embodying the ideal fulfilment of a rational urge to harmonize all one's projects. It is also supposed to provide a special grounding for the idea that a more fundamental form of regret is directed to deliberative error than to mere mistake. The regret takes the form of self-reproach, and the idea is that we protect ourselves against reproaches from our future self if we act with deliberative rationality: "nothing can protect us from the ambiguities and limitations of our knowledge, or guarantee that we find the best alternative open to us. Acting with deliberative rationality can only ensure that our conduct is above reproach, and that we are responsible to ourselves as one person over time."[8] These strains come together in Rawls' advocacy of ". . . the guiding principle that a rational individual is always to act so that he need never blame himself no matter how things finally transpire."[9]

Rawls seems to regard this injunction as, in a sense, formal, and as not determining how risky or conservative a strategy the agent should adopt; but it is worth remarking that if any grounding for self-reproach about deliberative error is to be found in the notion of the recriminations of one's later self, the injunction will in fact have to be taken in a more materially cautious sense. For the grounding relies on an analogy with the responsibility to other persons: I am a trustee for my own future. If this has any force at all, it is hard to see why it does not extend to my being required, like any other trustee, to adopt a cautious strategy with the entrusted goods—which are, in this case, almost everything I have.

However that may be, the model that gives rise to the injunction is false. Apart from other difficulties,[10] it implicitly ignores the obvious fact that what one does and the sort of life one leads condition one's later desires and judgments: the standpoint of that retrospective judge who will be my later

self will be the product of my earlier choices. So there is no set of preferences both fixed and relevant, relative to which the various fillings of my life-space can be compared; if the fillings are to be evaluated by reference to what I variously, in them, want, the relevant preferences are not fixed, while if they are to be evaluated by what I now (for instance) want, this will give a fixed set of preferences, but one which is not necessarily relevant. The recourse from this within the life-space model is to assume (as Utilitarianism does) that there is some currency of satisfactions, in terms of which it is possible to compare quite neutrally the value of one set of preferences together with their fulfilments, as against a quite different set of preferences together with their fulfilments. But there is no reason to suppose that there is any such currency, nor (still less) that the idea of practical rationality should implicitly presuppose it.

If there is no such currency, then we can only to a limited extent abstract from the projects and preferences we actually have, and cannot in principle gain a standpoint from which the alternative fillings of our life-rectangle could be compared without prejudice. The perspective of deliberative choice on one's life is constitutively *from here*. Correspondingly the perspective of assessment with greater knowledge is necessarily *from there*, and not only can I not guarantee how factually it will then be, but I cannot ultimately guarantee from what standpoint of assessment my major and most fundamental regrets will be.

For many decisions which are part of the agent's ongoing activity (the "normal science", so to speak, of the moral life) we can see why it is that the presence or absence of regrets is more basically conditioned by the retrospective view of the deliberative processes, than by the particular outcomes. One-self and one's viewpoint are more basically identified with the dispositions of rational deliberation, applicable to an ongoing series of decisions, than they are with the particular projects which succeed or fail on those occasions. But there are certain other decisions, as on the cases we are considering, which are not like this. There is indeed some room for the presence and subsequent assessment of deliberative rationality: the agents in our cases might well not be taken as seriously as they would

otherwise if they did not, to the limited extent which the situation permits, take as rational thought as they can about the realities of their situation. But this is not the aspect under which they will primarily look back on it, nor is it as a contribution to a series of deliberative situations that it will have its importance for them; though they will learn from it, it will not be in that way. In these cases, the project in the interests of which the decision is made is one with which the agent is identified in such a way that if it succeeds, his standpoint of assessment will be from a life which then derives an important part of its significance for him from that very fact; while if he fails, it can, necessarily, have no such significance in his life. If he succeeds, it cannot be that while welcoming the outcome he more basically regrets the decision; while if he fails, his standpoint will be of one for whom the ground project of the decision has proved worthless, and this (under the simplifying assumption that other adequate projects are not generated in the process) must leave him with the most basic regrets. So if he fails, his most basic regrets will attach to his decision, and if he succeeds, they cannot. That is the sense in which his decision can be justified, for him, by success.

On this account, it is clear that the type of decisions we are concerned with is not merely very risky ones, or even very risky ones with a substantial outcome. The outcome has to be substantial in a special way—in a way which importantly conditions the agent's sense of what is significant in his life, and hence his standpoint of retrospective assessment. It follows from this that they are, indeed, risky, and in a way which helps to explain the importance for such projects of the difference between extrinsic and intrinsic failure. With an intrinsic failure, the project which generated the decision is revealed as an empty thing, incapable of grounding the agent's life. With extrinsic failure, it is not so revealed, and while he must acknowledge that it has failed, nevertheless it has not been discredited, and may, perhaps in the form of some new aspiration, contribute to making sense of what is left. In his retrospective thought, and its allocation of basic regret, he cannot in the fullest sense identify with his decision, and so does not find himself justified; but he is not totally alienated from it, either, cannot just see it as a disastrous error, and so does not find himself unjustified.

This structure of retrospective understanding can occur without the concern introduced by the interests of others, which is central to our cases; but that concern is likely to be present in such decisions, and certainly it contributes importantly to their nature when it is present. The risks taken by our agents are taken in part with others' goods. The risks are taken also with their own, which increases our respect for them. But for themselves, they have a chance of winning, while the others do not; worse off than those served by the gambling trustee, the others' loss is settled from the start. There is no ground, whatever happens, for demanding that they drop their resentment. If they are eventually going to feel better towards him, it will not be through having received an answer to their complaints—nor, far from it, need it be because the agent is successful. They are not recompensed by the agent's success—or only if they are prepared to be.

But what about the rest of us? Here, for the first time, it is worth mentioning a difference between our cases, that if Gauguin's project succeeds, it could yield a good for the world as Anna's success could not. There is no reason why those who suffer from Gauguin's decision should be impressed by this fact, and there are several reasons (one of which we touched on earlier, in the matter of moral justification) why Gauguin should not. Nor should we be overimpressed by the difference, in considering what can be learned from such cases. But eventually the spectator has to consider the fact that he has reason to be glad that Gauguin succeeded, and hence that he tried. At the very least, this may stand as an emblem for cases in which we are glad. Perhaps fewer of us than is pretended care about the existence of Gauguin's paintings, but we are supposed to care, which gives an opportunity for reflection to start out and work towards the cases where we really care, where we salute the project. The fact is that if we believe in any other values at all, then it is likely that at some point we shall have reason to be glad that moral values (taken here in the simple sense of a concern for others' rights and interests) have been treated as one value among others, and not as unquestionably supreme. Real supremacy of the moral would imply its ubiquity. Like Spinoza's substance, if it were to be genuinely unconditioned, there would have to be nothing to condition it.

There is a public dimension of appreciation for such cases: how Gauguin stands with us (taking him emblematically as one whose project is saluted); whether we are, taking it all together, glad that he did it; depends on his success. That question, moreover, whether we are, taking it all together, glad, is the question we should take seriously. The various dichotomies which can be brought in to break up that question —such as moral *v.* non-moral, or agent *v.* act, or act *v.* outcome —often only help to evade the basic and connected questions of what one really wants the world to be like and what human dispositions are involved in its being like that.

These questions for the spectator we will leave; they would arise, as we noticed at the beginning, even if the agent had no concern for others' interests at all. But assuming (as we have throughout) that he has such a concern, then for him success makes a special kind of difference. It runs against the widely held view mentioned before, that moral regret is ultimate, and ultimate regret is immune to luck. If he fails, above all if he intrinsically fails, nothing is left except the cost to others for which (we are supposing) he in any case feels regret. In success, it must be dishonest or confused of him to regard that regret as his most basic feeling about the situation; if it were, he would at the most basic level wish that he had acted otherwise. In failure, that regret can consistently be part of his most basic feelings about what he has done. This is one way—only one of many—in which an agent's moral view of his life can depend on luck.

NOTES

[1] The question centres on the rôle of the Categorical Imperative. On the major issue here, I agree with what I take to be the substance of Philippa Foot's position ("Morality as a System of Hypothetical Imperatives", *Phil. Rev.* 1972; and her reply to Frankena, *Philosophy* 1975), but not at all with her way of putting it. In so far as there is a clear distinction between categorical and hypothetical imperatives, and in so far as morality consists of imperatives, it consists of categorical imperatives. The point is that the fact that an imperative is (in this sense) categorical provides no reason at all for obeying it. Nor need Kant think it does: the authority of the Categorical Imperative is supposed (mysteriously enough) to derive not just from its being (in this sense) categorical, but from its being categorical and self-addressed by the agent as a rational being.

[2] That acceptance is central to tragedy, something which presses the question of how we want to think about these things. When Oedipus says "I did not

do it" (Sophocles *OC* 539) he speaks as one whose exile proclaims that he did do it, and to persons who treat him as quite special because he did. Could we have, and do we want, a concept of agency by which what Oedipus said would be simply true, and by which he would be seeing things rightly if for him it was straight off as though he had no part in it? (These questions have little to do with how the law should be: punishment and public amends are a different matter.)

[3] For some discussion of this see "Ethical Consistency", in *Problems of the Self* (Cambridge 1973), pp. 166-186.

[4] A useful outline of such considerations is in D. A. J. Richards, *A Theory of Reasons for Action* (Oxford 1971), ch. 3.

[5] *Op. cit.* pp. 70-71, and *cf.* ch. 13.

[6] The notion of treating cases together, as opposed to treating them separately but in the light of experience, applies not only to deliberation which yields in advance a conjunctive resolution of a number of cases, but also to deliberation which yields hypothetical conclusions to the effect that a later case will receive a certain treatment if an earlier case turns out in a certain way: as in a staking system.

[7] John Rawls, *A Theory of Justice* (Oxford, 1972), esp. ch VII; Thomas Nagel, *The Possibility of Altruism* (Oxford, 1970).

[8] Rawls, pp. 422-423.

[9] p. 422.

[10] It ignores also the very basic fact that the size of the rectangle is up to me: I have said something about this in "Persons, Character and Morality", in Amélie Rorty, ed., *The Identity of Persons*, (California UP, forthcoming).

MORAL LUCK

B. A. O. Williams and T. Nagel

II—T. Nagel

Williams sidesteps the fascinating question raised in his paper.[1] He does not defend the possibility of moral luck against Kantian doubts, but instead redescribes the case which seems to be his strongest candidate in terms which have nothing to do with moral judgment. Gauguin's talent as a painter may be a matter of luck, but it does not, according to Williams, warrant the retrospective judgment that his desertion of his family is morally acceptable. In fact, it does not warrant any judgment about his prior decision that pretends to objective validity for everyone, or even to timeless validity for him. According to Williams, the effect of the fortunate outcome on Gauguin's attitude to his earlier choice will be merely to make him not regret, at the most basic level, having made it. He will not regret it because it has resulted in a success which forms the centre of his life. This attitude can hardly be called a judgment at all, let alone a moral judgment. Williams says Gauguin cannot use it to justify himself to others. It does not even imply the truth of an hypothetical judgment 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." And if the rest of us are glad that Gauguin left his family, Williams says that this is because we do not always give priority to moral values.

The importance of luck in human life is no surprise, even in respect of those matters about which we feel most deeply glad or regretful. It is the place of luck in ethics that is puzzling. Williams misdescribes his result in the closing paragraph of the paper: he has argued not that an agent's moral view of his life can depend on luck but that ultimate regret is not immune to luck because ultimate regret need not be moral. This is consonant with his tendency, here and in other recent writings,[2] to reject the impersonal claims of morality in favour of more personal desires and projects. Even if Williams has

successfully explained away the appearance of moral luck in the case of Gauguin, however, the explanation applies only to a narrow range of phenomena and leaves most of the area untouched. Williams acknowledges that he has dealt with only one type of case, but I do not believe these cases can be treated in isolation from the larger problem.

Why is there a problem? Not because morality seems too basic to be subject to luck. Some very important non-moral assessments of people deal with what is not their fault. We deplore madness or leprosy in ourselves and others, we rejoice in beauty or talent, but these, though very basic, are not moral judgments. If we ask ourselves why, the natural explanation is that these attributes are not the responsibility of their possessors, they are merely good or back luck. 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. This proposition uses an unanalysed concept of moral assessment that is presumably logically independent of the idea of control—otherwise the problem could not arise. Such a judgment is different from the evaluation of something as a good or bad thing, or state of affairs. The latter may be present in addition to moral judgment, 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 judgment 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.

However, if the condition of control is consistently applied, it threatens to erode most of the moral assessments we find it natural to make. For in various ways, to be discovered, the things for which people are morally judged are not under their control, or are determined to some extent 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 judgments intact.

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 judgments, without yielding the unacceptable conclusion derived from the broader condition, that most or all ordinary moral judgments 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 judgment 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 judgment 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.[8] 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 depend 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 judgments.

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 mentioned by Williams at the beginning of his paper—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. Williams' discussion is confined to the last category, but 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—the type of case Williams examines. We may note that the category includes a range of examples, from the truck driver who accidentally runs over a child to Gauguin and beyond. The driver, if he is entirely without fault, will feel

terrible about his rôle in the event, but will not have to reproach himself. Therefore this example of what Williams calls agent-regret 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 its 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, as Williams says, 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 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.[4]

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 judgments, 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 judgment.

I have said that Williams does not defend the view that these are instances of moral luck. The fact that Gauguin will or will not feel basic regret over his decision is a separate matter, and does nothing to explain the influence of actual results on culpability or esteem in those unquestionably ethical cases ranging from negligence through political choice. In such cases 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 don't 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 judgment does depend on the outcome, it is objective and timeless and not dependent on a change of standpoint produced by success or failure. The judgment after the fact follows from an hypothetical judgment 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 judgment 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 judgment 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 judgments.

> "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."[5]

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.[6]

Nevertheless the tendency to cut down the scope of moral assessment is 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 what Williams calls 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 judgment 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 conceited, but *behave* perfectly by a monumental effort of will. To possess these vices is to be unable to help having certain feelings under certain circum-

stances, 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 be in principle 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.[7] 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 shouldn't be like that, not that it's unfortunate that you are.

Nevertheless, Kant's conclusion remains intuitively unacceptable. We may be persuaded that these moral judgments 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.[8]

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 didn't 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.[9]

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.)

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 paper 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 judgment, 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—thought 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 judgments 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.[10] It is essentially what Williams means when he says, above,

> "One's history as an agent is a web in which anything that is the product of the will is surrounded and held up and partly formed by things that are not, in such a way that reflection can go only in one of two directions: either in the direction of saying that responsible agency is a fairly superficial concept, which has a limited use in harmonizing what happens, or else that it is not a superficial concept, but that it cannot ultimately be purified—if one attaches importance to the sense of what one is in terms of what one has done and what in the world one is responsible for, one must accept much that makes its claim on that sense solely in virtue of its being actual."

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 judgment 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 judgment is threatened with dissolution by the absorption of its acts and impulses into the class of events. Moral judgment of a person is judgment 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 unfortunote 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 didn't exist, or hadn't 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. Williams is right to point out the important difference between agent-regret and regret about misfortunes from which one is detached, but he does not emphasise the corresponding distinction in our attitudes toward others, which comes from the extension to them of external agent-centred evaluations corresponding to the agent-regret that they can feel about themselves. This causes him to miss the truly moral character of such judgments, which can be made not only by the agent himself, though they involve the agent's point of view.

There is a close connexion between our feelings about ourselves and our feelings about others. Guilt 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 what Williams calls 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 judgment—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.[11]

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 its presence is—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.

NOTES

[1] "Moral Luck", *Aristotelian Society Supplementary Volume* 1976.

[2] "Egoism and Altruism", in *Problems of the Self* (Cambridge, 1973); "Persons, Character, and Morality", in A. Rorty, ed., *The Identities of Persons* (Berkeley, Calif., forthcoming).

[3] See Thompson Clarke, "The Legacy of Skepticism", *Journal of Philosophy* LXIX (1972) 754-769.

[4] For a fascinating but morally repellent discussion of the topic of justification by history, see Maurice Merleau-Ponty, *Humanism and Terror* (Beacon Press, Boston: 1969).

[5] Part II, Section III, Introduction, paragraph 5.

[6] "Problematic Responsibility in Law and Morals", in Joel Feinberg, *Doing and Deserving* (Princeton, 1970).

[7] ". . . 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?" *Foundations of the Metaphysics of Morals*, Akademie edition p. 398.

[8] *Cf.* Thomas Gray, "Elegy Written in a country churchyard":

"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 T. Nagel, "War and Massacre", *Philosophy and Public Affairs* Vol. 1 No. 2 (Winter 1972); and B. Williams, "Ethical Consistency", *PASS* XXXIX (1965), also in *Problems of the Self*.

[9] 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 connexion 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 'sixties.

[10] 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 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.

[11] See P. F. Strawson's discussion of the conflict between the objective attitude and personal reaction attitudes in "Freedom and Resentment", *Proceedings of the British Academy*, 1962, reprinted in Strawson, ed., *Studies in the Philosophy of Thought and Action* (London, O.U.P., 1968), and in Strawson, *Freedom and Resentment and other essays* (London, Methuen, 1974).