

## Moral Luck\*

*Bernard Williams*

There has been a strain of philosophical thought which identifies the end of life as happiness, happiness as reflective tranquillity, and tranquillity as the product of self-sufficiency—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.<sup>1</sup> Just as, in the realm of character, it is motive that counts, not style, or powers, or endowment, so in action it is not changes actually effected in the world, but intention. With these considerations there is sup-

---

\*From *Moral Luck* (Cambridge: Cambridge University Press, 1981), ch. 2.

posed to disappear even that constitutive luck from which the ancient sages were happy to benefit. The capacity for moral agency is supposedly present to any rational agent whatsoever, 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.

It can offer that solace, however, only if something more is granted. Even if moral value were radically unconditioned by luck, that would not be very significant if moral value were merely one kind of value among others. Rather, moral value has to possess 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.

Any conception of 'moral luck', on this view, is radically incoherent. The phrase indeed sounds strange. This is because the Kantian conception embodies, in a very pure form, something which is basic to our ideas of morality. Yet the aim of making morality immune to luck is bound to be disappointed. The form of this point which is most familiar, from discussion of freewill, is that the dispositions of morality, however far back they are placed in the direction of motive and intention, are as 'conditioned' as anything else. However, the bitter truth (I take it to be both) that morality is subject, after all, to constitutive luck is not what I am going to discuss. The Kantian conception links, and affects, a range of notions: morality, rationality, justification, and ultimate or supreme value. The linkage between those notions, under Kantian conception, has a number of consequences for the agent's reflective assessment of his own actions—for instance, that, at the ultimate and most important level, it cannot be a matter of luck whether he was justified in doing what he did.

It is this area that I want to consider. I shall in fact say very little until the end about the moral, concentrating rather on ideas

of rational justification. This is the right place to start, I believe, since almost everyone has some commitment to ideas of this kind about rationality and justification, while they may be disposed to think, so far as morality is concerned, that all that is in question is the pure Kantian conception, and that conception merely represents an obsessional exaggeration. But it is not merely that, nor is the Kantian attempt to escape luck an arbitrary enterprise. The attempt is so intimate to our notion of morality, in fact, that its failure may rather make us consider whether we should not give up that notion altogether.

I shall use the notion of 'luck' generously, undefinedly, but, I think, comprehensibly. 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 view 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 me here, 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.

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 things 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 clearer what will count for him as eventual success in his project—at least, some possible outcomes will be clear examples of success (which does not have to be the same thing as 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 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. If he succeeds, he does have a basis for that thought.

As I have already indicated, I will leave to the end the question of how such notions of justification fit in with distinctively moral ideas. One should be warned already, however, that, even if Gauguin can be ultimately justified, that need not provide him with any way of justifying himself to others, or at least to all others. Thus he may have no way of bringing it about that those who suffer from his decision will have no justified ground of reproach. Even if he succeeds, he will not acquire a right that they accept what he has to say: 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 thought to be essential to rationality and to the notion of justification itself,

which is that one should be in a position 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. I do not want, at this stage of the argument, to lay much weight on the notion of morality, but it may help to throw some light on the matter of prior justification if we bring in briefly the narrower question whether there could be a prior justification for Gauguin's choice in terms of moral rules.

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 doubts about its content, the 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.

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 towards 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 there is something more interesting than that kind of difficulty. The Utilitarian perspective, not uniquely but clearly, will miss a very important dimension of such cases, 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 retrospectively, and while there is nothing 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. One is a distinction between luck intrinsic to the project, and luck extrinsic to it; the other is a 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 projects' 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 on 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 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 from her son. She might 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. 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. If Anna had then committed suicide, her thought might 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 such thoughts that I want 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 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.



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 himself 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 regret felt by a spectator comes 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 symbolize 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 always 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. It 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.

In these 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 this 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.<sup>2</sup>

The examples of Gauguin and Anna Karenina 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 agents' judgments on their decisions by what, beyond their 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 the 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 changes 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.<sup>3</sup> 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 process, 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.<sup>4</sup>

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 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 error, oversight, 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,<sup>5</sup> the agent's resolutions for his future deliberations. His regrets about his deliberations express themselves as resolves 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 pay-off) 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 makes 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.<sup>6</sup>

Insofar 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 if one knew enough, virtually 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 far more ambitious than 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.<sup>7</sup> 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.'<sup>8</sup> 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'.<sup>9</sup>



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. 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,<sup>10</sup> 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 that 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 fulfillments, as against a quite different set of preferences together with their fulfillments. But there is no reason to suppose that there is any such currency, nor 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. Oneself 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 in 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 that the situation permits, take such 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 that then derives an important part of its significance for him from that very fact; 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. 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 decisions we are concerned with are 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, never-

theless 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.

What is the relation of all this, finally, to morality? Does it have any very direct relation? Thomas Nagel,<sup>11</sup> who agrees that morality is deeply and disquietingly subject to luck, denies that an example such as Gauguin's shows that to be so—rather, it shows that Gauguin's most basic retrospective feelings do not have to be moral.

One reason that Nagel gives for this understanding of the matter is that (as I suggested earlier) Gauguin may not be able to justify himself to others, in the sense that they will have no justified grievance. However, this consideration just in itself will not carry great weight unless one makes a strong assumption about the nature of ethical consistency, to the effect that, if someone had acted justifiably from a moral point of view, then no-one can justifiably complain, from that point of view, of his so acting. But this as a general requirement is unrealistically strong, as can be seen from political cases,<sup>12</sup> for instance, in which we can have reason to approve of the outcome, and of the agent's choice to produce that outcome, and of his being an agent who is able to make that choice, while conscious that there has been a 'moral cost'. It is not reasonable, in such a case, to expect those particular people who have been cheated, used or injured to approve of the agent's action, nor should they be subjected to the patronising thought that, while their complaints are not justified in terms of the whole picture, they are too closely involved to be able to see that truth. Their complaints are, indeed, justified, and they may quite properly refuse to accept the agent's justification which the rest of us may properly accept. The idea that there has been a moral cost itself implies that something bad has been done, and, very often, that someone has been wronged, and if the people who have been wronged do not accept the justification, then no-one can demand that they should. It is for them to decide how far they are prepared to adopt the perspective within which the justification counts. This is just one of the ways—the distancing of time is another—in which, if the moral sentiments are to be part of life as it is actually experienced, they cannot be modelled on a view of

the world in which every happening and every person is at the same distance.

Our cases are admittedly different from the case of the politician. There, the justifying conditions relate to issues of what we want effected, what system of government we want, what persons we want to work within that system, and those wants may themselves be shaped by what are, in an everyday sense, moral considerations. With the agents in our examples, it is not the same, and there is, moreover, a difference between the examples themselves. If Gauguin's project succeeds, it can yield a good for the world as Anna's success could not. The moral spectator has to consider the fact that he has reason to be glad that Gauguin succeeded, and hence that he tried—or if a particular spectator finds that he has no disposition to be grateful for Gauguin's paintings, or for paintings, then there will be some other case.

It may be said that this merely represents our gratitude that morality does not always prevail—that moral values have been treated as one value among others, not as unquestionably supreme. I think that that misdescribes our relation to *this* Gauguin, at least, but it is important also to bear in mind the grounds, the scope and the significance of that gratitude, which I mentioned earlier, for the limitations of morality. If the moral were really supreme, it would have to be ubiquitous: like Spinoza's substance, if it were to be genuinely unconditioned, there would have to be nothing to condition it.

That is a demand which, only too familiarly, can extend itself among the feelings. The ultimate justice which the Kantian outlook so compellingly demands requires morality, as immune to luck, to be supreme, and while that does not formally require that there be no other sentiments or attachments, in fact it can, like the Robespierrean government to which Heine compared the Kantian system in general, steadily grow to require a wider conformity of the sentiments. Justice requires not merely that *something I am* should be beyond luck, but that *what I most fundamentally am* should be so, and, in the light of that, admiration or liking or even enjoyment of the happy manifestations of luck can seem to be treachery to moral worth. That guilty leveling of the sentiments can occur even if one recognizes, as Kant recognized, that there are some things that one is responsible for, and others for which one is not. The final destruction occurs when the Kantian sense of justice is joined to a Utilitarian con-

ception of negative responsibility, and one is left, at any level of importance, only with purely moral motivations and no limit to their application. There is, at the end of that, no life of one's own, except perhaps for some small area, hygienically allotted, of meaningless privacy.

Because that is a genuine pathology of the moral life, the limitation of the moral is itself something morally important. But to regard Gauguin's decision simply as a welcome incursion of the amoral is anyway too limited. It will be adequate only if he is the amoral Gauguin we put aside at the beginning. If he is not, then he is himself open to regrets for what he has done to others, and, if he fails, then those regrets are not only all that he has, but, as I have tried to explain, he no longer even has the perspective within which something else could have been laid against them. That can make a difference to the moral spectator. While he may admire the amoral Gauguin's achievements, and indeed admire him, this other Gauguin is someone who shares the same world of moral concerns. The risk these agents run is a risk within morality, a risk which amoral versions of these agents would not run at all.

The fact that these agents' justifications, if they acquired them, would not properly silence all complaints, does not itself lead to the conclusion that they are not moral justifications. However, perhaps we should, all the same, accept that conclusion. Their moral luck, we should then say, does not lie in acquiring a moral justification. It lies rather in the relation of their life, and of their justification or lack of it, to morality. That relation has to be seen in the first instance in their perspective, one in which, if they fail, there is simply regret. But their life is recognizably part of moral life, and it has a significance for us as well.

There is now, however, a pressing question—how much is being done by the concept of the moral, and how much *by this stage of the argument* does it matter what happens to it? In reminding ourselves of the significance of luck to the moral life—whether it is constitutive luck, or that which affects the relations of one's decisions to morality, or that which affects merely what one will turn out to have done—we essentially use the concept, because we are working out in reflection from central applications of the concept to question what may be a basic motive for using it all:<sup>13</sup> the motive of establishing a dimension of decision and assessment which can hope to transcend luck. Once that



motive is understood and questioned, it has to be asked once more what the concept is for, and, by the same token, how many other features of it can be taken for granted.

Scepticism about the freedom of morality from luck cannot leave the concept of morality where it was, any more than it can remain undisturbed by skepticism about the very closely related image we have of there being a moral order, within which our actions have a significance which may not be accorded to them by mere social recognition. These forms of skepticism will leave us with a concept of morality, but one less important, certainly, than ours is usually taken to be; and that will not be ours, since one thing that is particularly important about ours is how important it is taken to be.

### Notes

1. Kant's own account of this centrally involves the role of the Categorical Imperative. On that issue, 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 itself presses the question of how we want to think about these things. When Oedipus says 'I did not do it' (Sophocles *Oedipus at Colonus* 539) he speaks as one whose exile and blindness proclaim 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. 166–86.



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. 70–1, and cf. ch. 13 [*Moral Luck* (Cambridge: Cambridge University Press, 1981)]

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. 422–3.

9. P. 422.

10. It ignores also the very basic fact that the size of the rectangle is up to me; see Chapter 1 [*Moral Luck*].

11. In his contribution to the symposium for which this paper was originally written: *Proc. Arist. Soc. Supp. Vol. L* (1976), reprinted with revisions in his *Mortal Questions* (Cambridge, 1979) [see chapter 3]. I have benefited from Nagel's paper and from discussion with him. I entirely agree with him that the involvement of morality with luck is not something that can simply be accepted without calling our moral conceptions into question. That was part of my original point: I have tried to state it more directly in the present version of this paper. A difference between Nagel and myself is that I am more skeptical about our moral conceptions than he is.

12. See Chapter 4 [*Moral Luck*].

13. As Nagel points out, the situation resembles to some degree that with skepticism about knowledge. The same idea indeed seems to be involved in both cases: the knower is one whose belief is non-accidentally true (for discussion, see my *Descartes, the Project of Pure Enquiry* (Harmondsworth, 1978), pp. 37 seq). However, the path taken by skepticism from these similar starting points, and its eventual effectiveness, seem to be very different in the two cases.