

# Killing and Equality

JEFF McMAHAN

*University of Illinois at Urbana-Champaign*

Although the belief that killing is normally wrong is as universal and uncontroversial a moral belief as we are likely to find, no one, to my knowledge, has ever offered an account of why killing is wrong that even begins to do justice to the full range of common sense beliefs about the morality of killing. Yet such an account would be of considerable practical significance, since understanding why some killings are wrong should help us to determine the conditions in which killing is *not* wrong. For, in those cases in which the reasons why killing is wrong do not apply, killing may be permissible or, if there are positive reasons that favour it, even morally required.

In this paper I will begin to examine a range of issues that are critical to understanding the morality of killing. I should state at the outset my belief that matters are bound to be considerably more complicated than I am able to indicate here. My aims are modest and primarily exploratory. Whatever significance this paper may have will consist more in the questions it raises than in the answers it tentatively offers.

## I. THE HARM-BASED ACCOUNT

One approach to the morality of killing is to compare the killing of persons with the killing of animals. For it is uncontroversial that the killing of animals is less wrong than the killing of persons (in the sense that in general it takes much less to justify the killing of an animal than it does to justify the killing of a person) and, if we could determine *why* this is the case, this could illuminate the reasons why killing is in general wrong. Either some of the reasons why killing is wrong in the case of persons do not apply in the case of animals or the reasons that apply in the case of persons apply less strongly in the case of animals. Or both of these things could be true. If we could identify what is missing in the case of animals, this would reveal a substantial part of the explanation of why killing *people* is wrong.

One way of attempting to explain why the killing of an animal is normally less seriously wrong than the killing of a person is to appeal to the view that the wrongness of killing varies, other things being equal, with the degree of harm that killing causes to the victim. This view – the *Harm-Based Account* of the wrongness of killing – draws support from our sense that the wrongness of killing must be explicable

© Edinburgh University Press 1995

*Utilitas* Vol. 7, No. 1 May 1995

principally in terms of the fact that the victim ends up dead – that is, that the victim is harmed in the most serious possible way. (The Harm-Based Account has to be refined in order to account for the wrongness of attempted, unsuccessful killing, presumably by focusing on expected or intended rather than actual harm. Though important, this refinement need not detain us here.)

According to this view, the killing of an animal is less wrong because the animal suffers less harm in being killed. There are several types of harm that an act of killing may cause to its victim. One is the infliction of pain. This, however, does not significantly distinguish the killing of persons from the killing of animals. In general, the most serious harm that killing causes is, of course, the harm involved in death. It is mainly this harm that distinguishes the killing of animals from the killing of persons; for the harm that an animal suffers by dying is normally substantially less serious than that which a person suffers. To understand why this is so, it is necessary to distinguish several dimensions to the badness of death. Here I will merely summarize an account of the badness of death that I have developed in more detail elsewhere.<sup>1</sup>

The badness of death for an individual is proportional, other things being equal, to the net good that the individual's future life would have contained had he not died when and how he did. Because death can deprive an animal of a future life that would have been worth living, animals can be, and usually are, harmed by death. But, because the goods that their lives would otherwise have contained are markedly inferior in both quality and quantity to those characteristic of a normal human life, the harm they suffer in losing those goods is slight in comparison with the harm death normally involves for a person. Moreover, the harm that an individual suffers in being deprived by death of a certain future good is also affected by at least two other factors. First, it matters how closely psychologically connected the individual at the time of death ( $t_1$ ) would have been (in the absence of death) with himself in the future at the time that the good would have occurred within his life ( $t_2$ ). If the psychological connections between the individual at  $t_1$  and  $t_2$  would have been weak, then both the strength of the interest he has at  $t_1$  in having the good at  $t_2$  and the degree of harm he suffers at  $t_1$  in being deprived of that good have to be discounted for the diminished degree of psychological connectedness. A related but distinct consideration is whether and to what extent the individual at  $t_1$  actually desires to have the good at  $t_2$ . If he has

<sup>1</sup> Jeff McMahan, 'Death and the Value of Life', *Ethics*, ic (1988), 32–61; and 'Preferences, Death, and the Ethics of Killing', *Preferences*, ed. Christoph Fehige, *et al.*, Berlin and New York, forthcoming. An integrated presentation will appear in Jeff McMahan, *Killing at the Margins of Life*, New York, forthcoming.

no desire for the good, or if his desire for it is weak, then the harm he suffers in losing it must be correspondingly further discounted.

Because most animals are not self-conscious at all, while those that are have only a rudimentary form of self-consciousness, they do not have desires for their own future good and their lives are only very weakly psychologically unified over time. Thus the already comparatively slight harm they suffer in losing the goods of their future lives through death must be discounted even further both for diminished psychological connectedness and for the absence of desire for the goods of which they are deprived.

There are other dimensions to the badness of death that are absent altogether in the case of animals. I will mention just one. Many of our activities are onerous in themselves but nevertheless worthwhile because they are instrumental to the achievement of goals that we believe to be important. When death intervenes to prevent the realization of goals in which one has invested time and effort, the death may be bad not just because it precludes the future good of realizing the goals but also because it renders one's previous efforts and sacrifices vain or pointless. In short, death may affect the overall value of the life in two ways: by preventing the addition of further value to the life and by retroactively diminishing the value of the life one has already had. The latter type of effect is, however, not possible in the case of an animal, since the activities of animals do not derive their value from being goal-directed in the requisite way.

So, for these various reasons, the harm that an animal suffers in dying is substantially less than that normally suffered by a person. It is worth noting that the foregoing account also explains the common intuition that the death of a late-term human foetus or newborn human infant is normally less bad than the death of a child or adult but worse than the death of an animal.<sup>2</sup> Because the foetus or infant loses virtually the whole of a human life, the net good of which it is deprived is very great – far greater than that which any animal loses through death, and indeed greater than that which a normal adult loses. Yet because the foetus or infant would, were it to live, be only remotely psychologically connected with itself in the future, the harm it suffers in being deprived of its future must be steeply discounted for diminished psychological connectedness and absence of desire. And there are of course no negative retroactive effects.

There is one further type of harm that killing may cause to the victim: the harm involved in the violation of autonomy. Again, however, because animals are not autonomous in the relevant sense, they

<sup>2</sup> I say 'late-term' because I believe that the human foetus during the early stages of pregnancy cannot be harmed at all. This belief is defended in *Killing at the Margins of Life*.

are immune to being harmed in this way. So, in summary, killing is ordinarily substantially less harmful to an animal than it is to a person, primarily because animals cannot be harmed by death to anything approaching the extent to which a person can and secondarily because they cannot suffer the harm involved in the violation of autonomy. Consequently, the Harm-Based Account implies that, other things being equal, killing a person is always more seriously wrong than killing an animal. And it also offers a plausible explanation of why the killing of animals of certain types is worse than the killing of animals of other types. For, because animals vary considerably in their capacities for welfare, some animals can be harmed more by death than others. For example, because a dog's life is normally richer and more psychologically unified than a frog's, dogs are generally harmed to a greater extent by death. Therefore the Harm-Based Account implies that it is normally more wrong to kill a dog than it is to kill a frog.

The Harm-Based Account does not, of course, hold that the degree of harm caused to the victim is the only factor that determines the degree to which an act of killing is wrong. It holds only that the wrongness of killing varies with the degree of harm caused to the victim when other relevant factors are constant. It can therefore recognize that a wide variety of other factors can affect the moral status of an act of killing and that these factors can interact in complex ways. Relevant factors divide into several distinct types. Some have to do with the killer's mode of agency. It seems relevant, for example, that killing normally involves doing harm rather than merely allowing harm to occur. And it may also be relevant whether the act that results in the victim's death is intended to affect the victim in a certain way as a means of achieving the killer's ends or whether the effect on the victim is merely foreseen but unintended. Other factors, which I call *defeaters*, may nullify or otherwise diminish the normal moral significance of the harm that killing causes. Two such factors are moral guilt on the part of the victim and the victim's requesting or consenting to be killed. Defeaters have to be invoked, in my view, in any satisfactory explanation of the permissibility of suicide, voluntary euthanasia, killing in self-defence or war and, if it is permissible, killing as a mode of punishment. Third, special relations and their attendant forms of partiality may also affect the morality of killing. There are numerous other factors as well. This catalogue is meant to be merely suggestive, not exhaustive.

These various factors do not, however, explain what is fundamentally wrong with killing, though they may *contribute* to making killing wrong by exacerbating certain other considerations or, alternatively, as in the case of defeaters, they may nullify or weigh against the

factors that tend to make killing wrong. Because they are not fundamental, I will for the most part put these other factors aside for the purposes of this paper.

## II. TWO PROBLEMS OF EQUALITY

The idea that the wrongness of killing is a function of the degree of harm that killing causes to the victim raises two profound problems of equality. One of these is that, if we explain why the killing of animals is less wrong by appealing to the fact that their capacities are lower and thus that they are harmed to a lesser degree by being killed, then we will also, it seems, be committed to the conclusion that it is substantially less wrong intentionally to kill congenitally severely mentally impaired members of our own species whose cognitive and emotional capacities and potentials are comparable to those of certain non-human animals. This, I believe, is a problem to which there is no solution that will salvage all our intuitions. Some very deep intuitions will have to be repudiated if we are to achieve consistency without arbitrariness. The only acceptable solution is that severely retarded human beings and non-human animals with comparable cognitive and emotional capacities must for the most part be treated with the same degree of consideration.<sup>3</sup> While personal and species-based partiality may justify our giving the severely retarded some limited degree of priority, recognition of the rough moral equality of the severely retarded and animals with comparable capacities requires that we acknowledge greatly enhanced protections for animals while accepting that certain constraints that have hitherto governed our treatment of the profoundly retarded should be relaxed. The point of convergence is variable depending on the form of treatment. We should, for example, hold constant our view of the wrongness of causing pain to the severely retarded and thus recognize that causing pain to animals with comparable capacities is equally objectionable. But we should probably not assimilate our understanding of the morality of killing animals to the traditional view of the killing of the severely retarded. Rather, while we should recognize that the killing of many types of animal is considerably more objectionable than we have previously supposed, we should also acknowledge that some of the constraints that we have traditionally accepted as governing both the killing of the severely retarded and letting them die (for example, the constraint that forbids physicians to kill an anencephalic infant in order to use its organs for transplantation) are unreasonably stringent.

<sup>3</sup> I argue in detail for this conclusion in *Killing at the Margins of Life*, ch. II.

The problem of equality with which this paper will be concerned is different, though it too arises from a recognition that, because human beings have higher cognitive and emotional capacities than animals, their lives are normally better and their deaths correspondingly worse or more harmful. The problem is that all of the ways in which people are relevantly different from animals are also ways in which even normal adult human beings differ from one other. For every relevant capacity that we possess but that animals do not, or that we possess to a higher degree than any animal, some normal people possess that capacity to a higher degree than others.

Consider, for example, two people, whom we may call Bright and Dull. Bright has higher cognitive and emotional capacities than Dull. He is more intelligent and imaginative, capable of deeper personal relations, and so on. He is also, by temperament, a more efficient converter of resources into well-being. He is naturally cheerful and energetic, whereas Dull is lugubrious and depressive. Assume that the main differences in cognitive and emotional capacity and temperament are genetic in origin and thus to a considerable degree unalterable. Just as our capacities make possible higher levels of well-being than are possible for animals, so Bright's higher capacities allow him to achieve a higher level of well-being than is possible for Dull. Thus, since his future life would, other things being equal, be better, his death would be worse.

Next imagine that Bright and Dull both live in the same village. Lt. Calley and Charlie Company enter the village. One soldier wantonly shoots and kills Bright while another wantonly kills Dull. If the wrongness of killing varies with the degree of harm to the victim, then the killing of Bright is more seriously wrong than the killing of Dull, other things being equal. This is not, however, what most of us believe. The common intuition, which is embodied in the law, is instead that the two killings are *equally* wrong despite the difference in the degree of harm that each causes to its victim. The difference in the degree of harm appears to make *no* difference to the degree to which the killings are wrong.

There are of course other respects in which people's lives may differ in such a way as to make one person's death worse or more harmful than another's. Some factors of this sort are contingent, involving social circumstances or luck. For purposes of illustration, assume that wealth is closely correlated with well-being. Now consider two people, Rich and Poor. If Rich's wealth would make his future life considerably more worth living than Poor's and if other things are equal, then Rich would be harmed to a greater degree by dying and thus by being killed than Poor would. Yet we do not believe that, if one of Calley's men shoots Rich while another shoots Poor, the latter's act is less wrong

or morally objectionable. Again, the difference in the degree of harm caused to the victims appears not to affect the comparative wrongness of the two killings.

In addition to cognitive and emotional capacities, temperament, and social contingencies, another conspicuous factor that affects the badness of death among normal adult human beings is life expectancy, of which the principal determinant is age. Consider Young, who is 20 years old, and Old, who is 90. Again, if one of Calley's men shoots Young while another shoots Old, both killings seem equally wrong despite the substantial divergence between the amounts of harm they cause.

These are all instances in which there is a striking failure of correlation between the degree to which an act of killing is wrong and the degree of harm it causes to its victim. There are, of course, other types of case in which this correlation breaks down. In cases involving defeaters, there may be little objection to inflicting very great harm – for example, if a person's guilt causes him to deserve or to be liable to suffer a significant harm. In the cases just reviewed, however, the failure of correlation goes in the other direction; for these are cases in which the infliction of a *lesser* harm is condemned at a level appropriate to a greater harm. This is particularly salient in the case involving Young and Old. Imagine that Old had only a very short time to live in any case, that killing him deprived him of at most only a few weeks or months of life, perhaps at a diminished level of quality. Even then the killings of Young and Old seem equally wrong. Since the killing of Young is very seriously wrong, it follows that the killing of Old must be very wrong indeed.

### III. ATTEMPTS TO PRESERVE THE HARM-BASED ACCOUNT

There are two broad types of response to these cases in which the degree to which an act of killing is wrong seems unrelated to the degree of harm it causes to the victim. One is to reject the intuitions as errors of moral phenomenology. The other is to accept the intuitions and seek to determine what it is in these comparisons that is apparently nullifying the normal role that the degree of harm caused by an act has in determining the extent to which the act is *pro tanto* morally wrong. The first sort of response – rejecting the intuitions – requires either a wholesale rejection of the authority of moral intuition or else an explanation of how we have been misled in our intuitive responses to these specific examples. I cannot here address the general issue of the status of intuition but I will cite and reply to a few

suggested accounts of why our intuitions may be misguided in these cases. I will then turn to the alternative strategy of seeing what can be said on behalf of our intuitions.

One deflationary explanation of our intuitions in these cases is that they derive from a failure to distinguish between the comparative evaluation of the two acts and the evaluation of the two agents. It may be, in other words, that our intuition is really only that the two agents reveal themselves to be equally bad or vicious; but we then illegitimately and perhaps unconsciously extrapolate this conclusion, *mutatis mutandis*, to the comparative evaluation of the two acts themselves. Imagine, for the sake of comparison, that there are two thieves, both of whom commit robberies with the aim of enriching themselves. One actually succeeds in stealing an object worth \$1,000 while the other steals an object that he believes is worth \$1,000 but is in fact worth only a dollar. We may be inclined to condemn them equally, since their characters are equally reprehensible, but the latter's *act* is objectively less wrong than the former's.

There are several things that can be said in reply to this. One is simply that one's evaluations of the agents and one's evaluations of their acts are certainly distinguishable on reflection and, in my case at least, I find that, even after I have cautioned myself about the danger of conflation, I still believe that the two *acts* are equally wrong. Another response would be to embellish the case so that it is no longer true that the agents are equally bad. We might imagine, for example, that, while the soldier who kills Young does so with gusto and relish, the other kills Old with profound reluctance and aversion and only to avoid the ridicule to which he would otherwise be subjected by his more manly fellow soldiers. Given this adjustment to the details of the case, we now evaluate the agent who kills Old less harshly than the sadist who kills Young. *Whether or not we retain our tendency to conflate act-evaluation with agent-evaluation*, we should now, on this account, evaluate the act of killing Old as less seriously wrong. For the act causes less harm and the agent is less reprehensible. Speaking for myself, however, I continue to find the acts equally wrong.

A second explanation of how our intuition might be an illusion appeals to the fact that both acts are terribly wrong. Our reaction may be influenced by the fact that the difference between the relative degree of wrongness of the two acts may seem insignificant *in comparison with* the absolute magnitude of the wrongness of each. The difference tends to be obscured by the enormity of the two crimes. (One instance of this tendency emerged during a dispute in the 1970s about how many civilians had been killed in Kampuchea by the Khmer Rouge. Some commentators claimed that the controversy was in fact irrelevant, that it should make no difference to our assessment of the



Khmer Rouge or their action whether they killed 100,000 or twenty times that many.)

As a defence of the Harm-Based Account, this is inadequate, for two reasons. First, this explanation presupposes that the killing of Old is a very great crime. But, if the wrongness of killing is a function of the degree of harm that it causes to the victim, and if Old really would have died soon anyway, so that the harm he suffers is comparatively minor, then killing him *cannot* have been a great crime. The explanation, in short, presupposes a claim that the Harm-Based Account is committed to denying. Moreover, according to the Harm-Based Account, the difference in wrongness between the two acts must *in fact* be enormous, other things being equal, and therefore not easily obscured in the way that this explanation suggests. For the harm that Young suffers is the loss of many decades of life worth living, while that which Old suffers is the loss of only a few weeks or months. If the degree of wrongness were a function of the degree of harm caused, then the difference in relative wrongness would greatly overshadow the absolute degree of wrongness of killing Old – exactly the reverse of what this explanation suggests.

A third and final critique of the common intuition counters it with another intuition. Suppose that one were on the scene in the village and could prevent either one or the other of the murders but not both. It seems reasonable to suppose that one ought, other things being equal, to prevent the killing of Young. And this suggests that that act is more seriously wrong, since one ought to prevent the more serious of the two wrongs. It is, however, a mistake to assume that which of two acts one should prevent is determined by which would be more wrong. For the importance of preventing others from doing wrong lies primarily in the prevention of the *consequences* of wrongful action. In so far as we should be concerned to prevent wrongdoing independently of the prevention of its consequences, that concern should be primarily agent-relative: that is, we should be concerned to avoid *doing* wrong. The prevention of wrongdoing in itself is not an important reason for preferring one course of action to another.

To see this, suppose that one could either prevent a terrorist from detonating a bomb that would kill fifty innocent people or prevent the accidental explosion of a gas main that would kill fifty other innocent people *and* cause minor injuries and trauma to a few others. If one could not prevent both, and assuming that there would be no further relevant differences in consequences (for example, the terrorist will subsequently be captured and prevented from causing further harm whatever one chooses to do), it is not unreasonable to suppose that one ought to prevent the explosion of the gas main. If so, then while the prevention of wrongdoing may be of some significance independently

of the prevention of its consequences, it cannot matter much. Thus, although it is true that one ought to prevent the killing of Young rather than the killing of Old, this is not because the killing of Young is more wrong but only because it has worse consequences.

There may be other ways of attempting to undermine the common intuition in these cases but I cannot think of any that are more persuasive than those I have reviewed. It therefore seems warranted to see whether a convincing case can be mounted in defence of the common intuition that the two killings in each of these pairs of cases are equally wrong.

#### IV. THE TWO-TIERED ACCOUNT

The claim that all the killings are equally wrong can in fact be reconciled with the Harm-Based Account provided that one accepts, first, that the harm caused by a killing is proportional to the badness of death for the victim and, second, that all human deaths are equally bad. It is the latter claim, of course, that is most controversial; but it has its defenders, apparently including Paul Ramsey, who writes that ‘all our days and years are of equal worth whatever the consequence; death is not more a tragedy at one time than another’.<sup>4</sup> It is, however, unclear exactly what this means. Does the claim that all our days are of equal worth imply that no one can have a bad day? And even if there is a sense in which all our days are of equal worth, it does not follow that more days could not be better than fewer.

Perhaps there is an alternative interpretation of the idea that all deaths are equally bad. It is not implausible to suppose that the badness of death is proportional to the value of the life one thereby loses. (As I noted earlier, I believe this assumption must carry a *ceteris paribus* clause. But we may put that aside for the present.) Combine this assumption with the liberal egalitarian claim that all people’s lives are equally valuable – that no one’s life matters more or is more valuable than anyone else’s – and one gets the conclusion that all deaths are equally bad.

This view, however, rests on a confusion between two senses of the notion of the value of a life. In one sense, the value of a life is determined by the character of its contents. The value of a life is the value that the life has for the person who lives it; it is equivalent to the extent to which the life is worth living. In this sense, not all lives have equal value. Some people are more fortunate: their lives go better than those of most others. There is, however, another sense in which all lives, or at least the lives of all persons, might more plausibly be said

<sup>4</sup> Paul Ramsey, *Ethics at the Edges of Life*, New Haven, 1978, p. 191.

to have equal value. The value of a life in this sense is the value or worth of the individual or person whose life it is. This value is largely or completely unaffected by what happens within the life – by the character of its contents. It is instead determined by the nature of the subject of the life – by the particular properties and capacities that make that individual the kind of thing that he or she is. Let us reserve the phrase *the value of a life* for the first of these two notions and refer to the second as *the worth of the individual* or person whose life it is.<sup>5</sup>

Is the badness of death correlated with the value of the life or with the worth of the individual? If the value of death were inversely proportional to the worth of the individual, then, provided that all persons have positive worth, it would follow that death could never be good for a person. But this is implausible; in some cases, when the contents of a person's future life would be well past the point at which the life would cease to be worth living, it can be good for that person to die, irrespective of his or her inherent worth as a person. The badness of death must therefore be a function of the value of life in the first sense. The badness of death is, other things being equal, commensurate with the expected value of the contents of a person's projected future life. But the value of life in this sense varies from person to person. Thus all deaths are not equally bad and, if the wrongness of killing varies with the badness of death for the victim, all killings cannot be equally wrong, even if all other things (such as moral innocence) are equal.

One possibility that suggests itself at this point is that the wrongness of killing varies, *not* with the degree of harm caused to the victim, which is a function of the value of the victim's life, but instead with the *worth* of the victim. According to this view, the *badness of death* is correlative with the value of a life but the *wrongness of killing* is correlative with the worth of the victim. If all persons have equal worth, then all wrongful killings of persons must be equally wrong. This would explain why the intentional killings of Bright and Dull, Rich and Poor, and Young and Old are all equally wrong.

Call the view that the wrongness of killing varies, other things being equal, with the worth of the victim the *Worth-Based Account*. This account appears to have the same virtues that I earlier ascribed to the Harm-Based Account; that is, it too offers plausible explanations of why killing persons is more seriously wrong than killing animals and why killing animals of some types is more wrong than killing animals of other types. To see this, we need to say something, however minimal, about the bases of individual worth. What is it about us that makes us, as individuals, more important or valuable or worthy than animals? The answer, it seems, has to do with the fact that we possess a range

<sup>5</sup> A similar distinction is drawn in Tom Regan, *The Case for Animal Rights*, London, 1983, pp. 235–6.

of cognitive and emotional capacities that are substantially more highly developed than those of animals.<sup>6</sup> Frequently cited examples include capacities for rationality, autonomy, free choice, imagination and creativity, the use of language, deep personal and social relations, empathy, and so on. If the degree to which killing is wrong is a function of the worth of the victim, and if worth is determined by the possession of capacities such as these, then the reason why it is more seriously wrong to kill people than it is to kill animals is that people possess all of these capacities to a high degree while animals either do not possess them at all or possess them to a markedly lower degree. And because animals of some types possess more of these properties, or possess some of them to a higher degree, than animals of other types, it is more seriously wrong to kill animals of the former types than it is to kill animals of the latter types, other things being equal.

According to the Worth-Based Account, the degree to which an act of killing is wrong is not directly affected by the degree of harm caused to the victim. But there remains a contingent correlation between the two. For those whom it is most wrong to kill are, on this view, those with the highest worth, and those with the highest worth are those with the highest cognitive and emotional capacities. But higher cognitive and emotional capacities also make possible a higher level of well-being – indeed, they entail a higher capacity for well-being. Hence those individuals with the highest worth will also tend to have lives with the greatest value – that is, lives that are most worth living. They are therefore likely to be harmed to a greater degree by death, and thus by being killed, than those of lesser worth. But the correlation here is only rough; for even those with the highest capacities are not immune to bad luck. In cases in which a person's future life would be short or of a low level of quality despite his high capacities (or perhaps even *because* of them, since unusual sensitivity can also make one vulnerable to depths of misfortune that are not possible for those with lesser cognitive endowments), the Worth-Based Account and the Harm-Based Account could diverge quite considerably in their assessments of how wrong it would be to kill that person.

Another virtue of the Worth-Based Account is that it seems to offer an explanation of the common view that, while neither the quality of life nor the life expectancy of the victim affects the degree to which an act of killing is wrong, the *number* of victims that the act has *does*. For, while neither the quality of a person's life nor the expected quantity of his future life affects his worth as a person or, therefore, the wrongness of killing him, it is at least arguable that killing a greater number of

<sup>6</sup> Virtually the only alternative accounts are theological in nature. I will not pursue these here, though in *Killing at the Margins of Life* I consider and reject the claim that what distinguishes us from animals is that we, but not they, have souls.

people constitutes a greater offence against the worth of persons than killing a lesser number.

As it stands, however, the Worth-Based Account is also vulnerable to serious objections. I will mention two. First, it makes the wrongness of killing vary with the worth of the victim all the way down to the lowliest animal, ignoring the degree of harm caused to the victim. At the lower end of the scale, this seems a mistake. To me it seems worse, or more objectionable, to kill a dog in the flush of youth than to kill a dog of advanced age in a state of physical decrepitude. The difference in the degree of harm caused seems relevant. Second, even among normal adult human beings, some possess the relevant cognitive and emotional capacities to a higher degree than others. If worth is determined by these capacities, then some normal people have a higher worth than others. The Worth-Based Account thus implies that it is more seriously wrong to kill people the higher their cognitive and emotional capacities are. While it may imply that the killings of Young and Old and Rich and Poor are equally wrong, it may also judge that the killing of Dull is less wrong than the killing of Bright – an implication that offends strong egalitarian intuitions.

Each of these objections may be addressed by positing a threshold. The first objection can be answered by stipulating that there is a threshold along the scale that measures overall cognitive and emotional capacity. The wrongness of killing beings that fall below the threshold is to be explained primarily in terms of the harm that the killing causes to the victim. Below the threshold, the degree to which an act of killing is wrong thus varies, other things being equal, with the degree of harm caused. Above the threshold, by contrast, the wrongness of killing is to be explained primarily in terms of the violation of a requirement to respect the victim's worth as an individual. The degree to which the killing of an individual above the threshold is wrong is thus commensurate, other things being equal, with that individual's worth and is unaffected by how great a harm he suffers in being killed.

This revision represents a melding of the Harm-Based Account and the Worth-Based Account. In effect, the Worth-Based Account explains the morality of killing above the threshold while the Harm-Based Account explains morality of killing below it. The threshold, on this view, marks the point at which the morality of killing is divided between two spheres of morality. Below the threshold, killing is governed by what might be called *the morality of interests*: the dimension of morality concerned with consequences, interests, and benefits and harms. Above the threshold, acts of killing are governed by *the morality of respect*.

The idea that there is a distinct domain within morality in which

the central requirement is one of respect is an acknowledgement that one's relations with certain individuals must be governed by more than a concern for their interests or well-being. It is an acknowledgement that certain acts may be seriously wrong, whether or not they cause harm, because they fail to respect an individual for what he is, fail to recognize or respect his status as one's moral equal. On one interpretation, Kant held that the morality of respect is in fact coextensive with the *whole* of morality – hence his denial that there could be duties with respect to animals other than those that are owed to people but of which animals are the accidental beneficiaries (for example, a duty to avoid cruelty to animals lest it coarsen one's nature, making it likelier that one will be cruel to people). This seems to me a mistake. Our treatment of animals does lie within the scope of morality, though it falls entirely within the domain of the morality of interests. Relations with others whose capacities vault them above the threshold are then additionally subject to the constraints of the morality of respect. Those above the threshold are granted a further set of protections that are denied to animals.<sup>7</sup> The threshold marks the point at which individuals not only merit consideration of their interests but also have sufficient worth to command respect. Let us therefore call this threshold the *threshold of respect*. Killing a being above this threshold constitutes the most egregious possible violation of the requirement of respect for individual worth.

Because the morality of killing above the threshold of respect is governed by considerations from a different area of morality from those that govern the morality of killing below the threshold, the comparative wrongness of killings above and below the threshold may be obscure. It seems likely that there may be forms of incommensurability here: for example, it might be argued that no act of killing whose victim or victims fall below the threshold can be as seriously wrong as an act of killing whose victim or victims are above the threshold, if other things are equal. I will not pursue this speculation here. The essential point is that killings that violate the requirement of respect are of a different and higher order of seriousness from those that do not.

While this hybrid account of the morality of killing that results from postulating the threshold of respect avoids the first of the two objections to the Worth-Based Account, it remains vulnerable to the second. For it retains the assumption that an individual's worth varies with his or her cognitive and emotional capacities. Thus, even above the threshold of respect, people will have different levels of worth and

<sup>7</sup> This view has clear affinities with a view sketchily advanced by Nozick under the label 'utilitarianism for animals, Kantianism for people'. (Robert Nozick, *Anarchy, State, and Utopia*, Oxford, 1974, p. 39.)

killing people with higher worth will be more wrong, other things being equal, than killing those of lower worth. One might articulate this implication by saying that people command different degrees of respect depending on their level of cognitive and emotional capacity. This objectionably inegalitarian implication can, however, be avoided by the introduction of a further threshold. One can stipulate that, somewhere along the scale that measures overall cognitive and emotional capacity, there is a point at which individual worth ceases to fluctuate with variations in cognitive or emotional capacity. There is a certain level of cognitive and emotional capacity such that the corresponding level of worth cannot be exceeded. All individuals at or above that level of capacity have equal worth. Let us therefore refer to this point on the scale as the *threshold of equal worth*.

In order to capture our intuitions about the killing of persons and the killing of animals, the Worth-Based Account must incorporate both these thresholds. These significant revisions result in an account that is sufficiently different to count as a distinct theory. Let us call it the *Two-Tiered Account*. According to the Two-Tiered Account, there is a threshold of respect such that the wrongness of killing beings with capacities that place them below the threshold varies with the degree of harm they suffer in being killed, while the wrongness of killing beings above the threshold is unaffected by the degree of harm they suffer but instead consists in the violation of a requirement of respect for their worth as individuals. There is also a second threshold such that all those above it have equal worth despite variations in overall cognitive and emotional capacity. It is plausible, moreover, to suppose that these two thresholds coincide: that the threshold of respect and the threshold of equal worth are located at the same point on the scale that measures overall cognitive and emotional capacity. (If they do coincide, then for most purposes we can refer to them both by speaking simply of 'the threshold'.) Let us also stipulate – though this is independently plausible – that those with psychological capacities above the threshold merit the label 'person', while those below it do not. Given these assumptions, the Two-Tiered Account implies that, in those cases in which killing individuals above the threshold is wrong, it is wrong primarily because it violates a requirement of respect for the equal worth of persons. And, if other things (such as considerations of agency, number of victims, and so on) are equal, then acts of killing individuals above the threshold are equally wrong, irrespective of the degree of harm their victims suffer by dying.

#### V. WHAT DOES RESPECT REQUIRE?

If we are to recognize an area of morality in which the principal requirement is one of respect, we will obviously want to know what is

required of us in order to show respect for those to whom it is due. According to some accounts of the ethics of respect, one may proceed inferentially from the concept of respect together with a set of claims about the grounds or bases of respect to conclusions about what sorts of act are or are not compatible with respect for persons.<sup>8</sup> For various reasons, some of which will emerge later in the paper, I do not believe that this is a plausible or fruitful strategy. I believe, in fact, that we must proceed in exactly the opposite way, despite misgivings about the seemingly unsystematic nature of this method of inquiry. I believe, in short, that we must begin, as I have done, with a set of intuitions about equality, constraints on action, and so on that arise from sustained reflection about particular problems and cases and then seek to organize, make sense of, and in some cases revise these beliefs within a coherent framework of principles and concepts. Thus I have been led by consideration of our intuitive responses to such cases as those involving Bright and Dull, Rich and Poor, and Young and Old to speculate that the Kantian tradition in moral philosophy has been right that there is a requirement of respect for persons. This notion seems to illuminate our sense that, while the wrongness of killing animals is fully explicable in terms of the harm that this killing causes, the various killings of persons that we have considered are *equally* wrong despite the substantial differences in the amounts of harm they cause. But what else is involved in or required by respect for persons is something that we have to discover, not by deduction, but by further exploration of how and why some of our deepest and most compelling intuitions are resistant to explanation or accommodation within the morality of interests and consequences.

I want to explore the intuitive terrain a bit more, though my aims are very limited. What is interesting about the cases we have considered is not so much that the wrongness of killing seems to be importantly affected by something other than the harmful nature of the action, but that it seems to be entirely *unaffected* by the amount of harm caused. It is not the case, in other words, that considerations drawn from the morality of respect combine with those from the morality of interests to determine the moral status of the various killings. The two dimensions of morality do not supplement or complement one another; rather, the morality of interests seems to be altogether eclipsed or supplanted by the morality of respect. Considerations of interests and considerations of respect are not additive in these cases. If they were, then the killing of Young would be more wrong, since

<sup>8</sup> See William K. Frankena, 'The Ethics of Respect for Persons', *Philosophical Topics*, xiv (1986), pp. 160–4, and Alan Donagan, *The Theory of Morality*, Chicago, 1977, pp. 66–74.



it would involve an *equal* violation of respect and yet cause a *greater* amount of harm. It would, that is, be opposed by a greater number, or a stronger array, of factors or considerations. Yet this is not the case; the killings are equally wrong.

Why is it that considerations of interests or harm seem irrelevant to the morality of killing in these cases? It is certainly not because the morality of respect *always* displaces the morality of interests in cases involving persons. Consider yet another variant of the case of the two soldiers. One soldier wantonly inflicts ten minutes' worth of moderately severe pain on one innocent villager, while a second soldier inflicts an hour's worth of extreme agony on another. In neither case is there any permanent damage. Here the common belief is that the two acts are *not* equally wrong: the infliction of the greater harm is more seriously wrong. So the morality of interests is operative in this case. Yet surely both acts of inflicting pain also involve a failure of respect. The fact that intentionally inflicting pain on an innocent person normally *does* involve a failure of respect is one part of the explanation why it is normally more objectionable to inflict a certain amount of pain on a person than it is to inflict the same amount of pain on an animal. (The other elements in the explanation are that pain has greater opportunity costs in the case of a person and that pain normally has more severe and extensive ramifications throughout the victim's life in the case of a person – for example, psychic scarring, anticipation, anxieties about the significance of the pain, and so on – though in some instances these latter considerations may be partially offset by the fact that, as Judith Thomson puts it, 'an adult human being can ... think his or her way around the pain to what lies beyond it in the future'.)<sup>9</sup>

I am unable to offer an explanation of why some acts seem to be governed entirely by the morality of respect, so that the extent to which such an act causes harm appears to be irrelevant to its moral status. It is, however, possible to say something about the conditions that seem to distinguish these cases from others. The foregoing comparison between acts that cause death and those that cause pain suggests that the irrelevance of the degree of harm caused may be confined to cases involving *lethal* harms. Perhaps, as I suggested earlier, the *annihilation* of an individual whose existence and nature demand respect, against that individual's will, is the ultimate violation of the requirement of respect, so egregious that all other considerations are simply superseded. (Note, however, that I refer to lethal harms and not just killings. If one soldier were wantonly to allow Old to die, when he could easily have saved him, most of us would find this equally

<sup>9</sup> Judith Jarvis Thomson, *The Realm of Rights*, Cambridge, Mass., 1990, p. 292.

wrong as wantonly allowing Young to die, other things being equal.)<sup>10</sup>

Thus far I have speculated that there are two conditions that must obtain in order for it to be the case that the amount of harm that an act causes is irrelevant to the moral status of the act and therefore to the degree to which the act is wrong. One is that the victim must be above the threshold of respect; the other is that the act must be instrumental in bringing about the victim's death. There are, however, cases in which an act meets both of these conditions and yet it is not irrelevant to the moral status of the act how much harm it causes to its victim. Here is an example.<sup>11</sup> A train carrying hundreds of passengers is headed at high speed for the station when its brakes fail. If it crashes into the station, hundreds of people will be killed. Imagine that one is a bystander but that one has access to a lever that can be used to divert the train onto a branchline track. Through no fault of his own, however, Young is trapped on this track. Intuitively it seems permissible to divert the train from the mainline track even though that will involve killing an innocent person. Yet surely killing Young involves a failure of respect. Is this a case of conflict between the morality of interests and the morality of respect, in which the former takes precedence because the importance of numbers outweighs the requirement of respect for Young? Or is it a case of conflict within the morality of respect, in which letting the hundreds die would also constitute a failure of respect? Though I suspect that the answers are yes and no, respectively, I confess that I am uncertain how to answer these questions. It seems clear that the morality of respect is sensitive to considerations of agency. For example, it seems that the requirement of respect is less than decisive in its opposition to diverting the train because this does not involve using Young or his death as an intended means of saving the hundreds. But it is less clear what the morality of respect implies about the failure to divert the train, since this would not only not involve doing harm but would also not involve intentionally affecting the victims. Beyond this, I am uncertain what to say.

These questions are, however, really only preliminary to the point I wish to make, which emerges only when we add another feature to the

<sup>10</sup> Similarly, suppose that one could either save Old, who would then have only a very short time to live, or a large number of stray dogs. It is not implausible to suppose that there is no number of dogs such that one ought to save them in preference to Old. Yet when we turn to non-lethal harms, Old ceases to have priority. If one could either prevent Old from suffering a certain amount of pain or prevent some number of dogs from each suffering the same amount of pain, it would take only a few dogs to make it more important to prevent their pain than to prevent Old's.

<sup>11</sup> This is a variant of the well-known trolley case first discussed by Philippa Foot in 'The Problem of Abortion and the Doctrine of Double Effect', reprinted in her *Virtues and Vices*, Los Angeles, 1978, pp. 19–32.

case. Suppose that there are two branchline tracks with Young on one and Old trapped on the other. It remains permissible to divert the train; but now there is a choice between killings. Should the train be diverted so that it kills Young or so that it kills Old? It is not implausible to suppose that one ought to choose to kill Old. At a minimum, it seems that the fact that killing Old would cause less harm is a relevant consideration. If so, then there are cases in which the moral status of an act of killing an individual above the threshold of respect is affected by the degree of harm that the act causes to the victim. In this case, the fact that killing Old would cause him less harm than Young would suffer if he were killed instead makes it preferable to kill Old. What is the relevant difference between this case and the case of the two soldiers, in which the killings of Young and Old seem equally wrong and therefore morally indistinguishable? There are of course many differences but the significant difference seems to be that in this case conditions are such that killing an innocent person has become permissible. One might speculate that, in cases in which it has become permissible to kill an innocent person, even if this involves a failure of respect, but in which there is a choice among possible victims, the failure of respect is not aggravated or exacerbated if one chooses to kill that person among the possible victims who would suffer the least harm in being killed. In short, in order for it to be the case that the morality of an act of killing is unaffected by the magnitude of the harm that it causes to its victim, the act must be wrong all things considered.

Two clarificatory points are necessary. First, this presupposes that the permissibility (or wrongness) of the act of killing can be established independently of how much harm the act causes to its victim. For permissibility is supposed to be a condition of the relevance of the degree of harm. Thus the degree of harm cannot be relevant in establishing that the condition obtains. Second, it might be objected that, if one ought to kill Old rather than Young in the case of the runaway train, then killing Young is, in the circumstances, wrong all things considered. Hence the degree of harm it causes to its victim should be irrelevant. Yet it is precisely because it would cause more harm that it seems to be wrong. To answer this objection we must relativize the relevance of certain considerations to the stage of deliberation to which they are appropriate. At the stage of deliberation when one is considering whether to divert the train so that it kills Young or Old, the levels of harm caused by the two killings may be relevant, for at that stage it has not yet been determined that killing Young would be wrong. All that is known at that stage is that either killing would be permissible if the other were not an option. Thus the degree of harm that killing Young would cause may be relevant to one's choice between the two killings.

It may be worth digressing briefly to note a further interesting point. I have suggested that there are certain conditions in which the moral status of an act seems unaffected by the degree of harm that it causes to its victim. These conditions are that the act should be instrumental in causing the death of an individual (the *lethality condition*), that the individual should be above the threshold of respect (the *threshold condition*), and that the act should be wrong, all things considered – that is, there are no countervailing conditions that override or neutralize the fact that the act violates the requirement of respect (the *wrongness condition*). If other things are equal, acts that meet these conditions are equally wrong, irrespective of the amount of harm they cause to their victims. If, however, one of the conditions is not met, then it seems that the degree of harm that an act causes becomes a relevant factor in determining the moral status of the act. And, interestingly, there are other considerations that appear to become relevant only in the absence of these three conditions. Among these is an ideal of equality that is quite different from the ideal of equal respect that seems to underlie our judgements in the various cases involving the two soldiers. The alternative ideal that I have in mind might be called the ideal of overall equality of well-being. It holds that, if individuals are not conspicuously unequal with respect to certain characteristics – notably virtue or merit – then it is best if they derive overall equal gains or benefits from life. Unless one person is notably more virtuous or meritorious than another, our sense of natural justice is offended if the one fares substantially better in life than the other – for example, is healthier or happier or lives a much longer life.

Recall Rich and Poor. Imagine that they are the same age, are equally virtuous and meritorious, and so on. Through no fault of his own, Poor has fared much worse in life than Rich. According to the ideal of overall equality of well-being, it is worse that Poor should die now, since death would limit his gains from life by more than it would limit Rich's. This seems a relevant consideration if we imagine a variant of the case of the runaway train in which Rich and Poor are each trapped on a different branchline track. While considerations of comparative harm favour diverting the train so that it will kill Poor (since his future life promises to be less good than Rich's), the ideal of overall equality favours killing Rich (since this would reduce the inequality between their overall lifelong levels of well-being while killing Poor would increase it).<sup>12</sup> Notice, however, that the fact that Poor's death is worse in terms of overall equality is irrelevant in the

<sup>12</sup> Some who think that we should be guided by considerations of comparative harm believe that this too is required by yet another ideal of equality: equal consideration of interests.

case of the two soldiers. It does nothing to make the killing of Poor more seriously wrong than the killing of Rich.

In short, the ideal of overall equality of well-being is relevant to the morality of killing when the wrongness condition fails to hold. It also seems relevant to the morality of action when the lethality condition is not met. But is it relevant to the morality of killing when the threshold condition is not met? Consider, for example, a further variant of the case of the runaway train in which a person is trapped on one branchline track while a dog is trapped on the other. Does the ideal of overall equality of well-being favour diverting the train so that it kills the person, since this would diminish inequality between his overall level of well-being and that of the dog? Clearly not; and this is not just because the dog would be devoid of virtue or merit (the person might score low in these respects as well). Rather, the scope of this ideal seems to be limited in the following way: it applies only among those with roughly comparable capacities for well-being. Our understanding of the good life for a dog is relativized to its cognitive and emotional capacities. If a dog does well relative to what its capacities permit, it has a good life. It is not unfair that its life is substantially less good overall than that of a person; nor is there any reason to try to compensate the dog for this or to narrow the gap in overall well-being between it and a person. (The ideal may reapply independently at different levels. It might imply that all those with a certain higher capacity for well-being should fare equally well as each other while all those with a certain lower capacity should also fare equally well as each other.)

One disturbing implication of this understanding of the scope of the ideal is that, if the ideal does not apply in comparisons between persons and non-human animals, then neither can it apply in comparisons between normal adult human beings and severely mentally retarded human beings whose cognitive and emotional capacities are comparable to those of a non-human animal. And, if the scope of other ideals of justice and equality are similarly constrained, then the severely mentally retarded may fall outside the scope of our principles of justice and equality altogether. This possibility raises profound problems, particularly for those who hold that natural assets, including physical and psychological capacities, fall within the scope of distributive principles. Some have argued, for example, that we have reason, as a matter of justice, to intervene genetically to correct severe physical disabilities where this is feasible.<sup>13</sup> And it is natural to suppose that this reasoning should be extended to cognitive disabilities

<sup>13</sup> See, for example, chs. 5 and 6 drafted by Allen Buchanan for a forthcoming book on ethics and the Human Genome Project, co-authored with Dan Brock, Norman Daniels, and Daniel Wikler.

as well. If, for example, it were possible to correct the genetic abnormality responsible for Down's Syndrome before the condition caused irreversible cognitive impairment, it might plausibly be argued that there would be a reason, grounded in considerations of justice or equality, to intervene. But (putting aside problems of personal identity) this assumes that an individual's cognitive capacity, which hitherto seems to have functioned as a boundary condition for the application of principles of justice and equality, may now be treated as a good whose distribution is governed by these same principles. Moreover, it is difficult to accept the assumption that there is a strong moral reason to enhance the cognitive capacities of severely retarded human beings without also being committed to the absurd conclusion that there is an equally strong moral reason to enhance the cognitive capacities of non-human animals. These, however, are problems best left for another occasion.

## VI. THE BASIS AND SCOPE OF RESPECT

I have suggested that there is a threshold such that the moral status of an act of killing an individual above the threshold is, unless there are extraordinary conditions that would make the act permissible, unaffected by the amount of harm the act would cause to the victim. I have also suggested that this threshold lies somewhere on the scale that measures overall cognitive and emotional capacity. Is this right? Is it really our psychological capacities that distinguish us so importantly from animals? If so, where on the scale does the threshold lie?

Some theorists, particularly in the tradition of Kantian moral philosophy (which is, in effect, the source of the notion of respect), believe that the answers to these questions can be derived from truths about the nature of morality. Some hold, for example, that the basis for respect is indeed a certain set of psychological capacities – namely, those that are necessary and sufficient for moral agency. Thus Kant argued that those who have dignity and worth, and who therefore command respect, are those with the capacity to guide or govern their lives in accordance with principles that they rationally and autonomously legislate for themselves. According to a similar account, the basis for respect is the possession of the capacities necessary for participation in contractual procedures that some claim are constitutive of the foundations of morality.<sup>14</sup>

<sup>14</sup> For one influential variant of this approach, see John Rawls, *A Theory of Justice*, Oxford, 1972, § 77.

Again, my own approach is altogether less systematic. I confess to being uncertain about the basis of respect. What I share with Kantians, liberal egalitarians, deontologists, and others is not an agreed understanding of the basis of respect but a conviction that there is something there to be understood. My suggested focus on cognitive and emotional capacities as the basis for respect arises from reflection about what it is that Young and Old share that makes each of them, as an individual, equally important or worthy as the other, as well as more important or valuable or worthy than any animal. I focus on these capacities because, again, this seems to illuminate – though I concede that the illumination is faint rather than blinding – our intuitive sense that, in the case of the two soldiers, the killings of Young and Old are equally wrong and that the comparative harm they suffer is irrelevant to this judgement. My predilection is to try to make progress by further exploring the intuitive terrain, especially around the edges where our doubts about the morality of respect are most acute. We may take it as a datum that the threshold of respect separates normal adult human beings from non-human animals; but where do fetuses, infants, and the severely retarded lie in relation to the threshold?

Understanding the basis of respect is profoundly important if only because it is necessary to understanding the morality of our treatment of these beings whose moral and metaphysical status is obscure. For the basis of respect determines the scope of the morality of respect. Consider the morality of abortion. One might attempt to defend the permissibility of abortion by appealing to the claim that the fetus suffers only a relatively modest harm in being killed, for reasons indicated in the brief account of the badness of death sketched earlier. Given the Harm-Based Account of the morality of killing, this claim would provide the basis for a powerful argument. But we have seen that the killing of Old in the case of the two soldiers is seriously wrong despite the fact that it causes comparatively little harm. If whatever it is that makes that killing so seriously wrong is also present in the case of abortion – if, that is, fetuses are above the threshold of respect – then abortion may be seriously wrong as well, for the same reason. If, however, fetuses fall below the threshold, then the claim that fetuses do not suffer a significant harm in dying can be integrated with the Two-Tiered Account to provide a powerful defence of abortion.

If it is right that the threshold of respect consists in the possession of an overall level of psychological capacity higher than that of at least most animals, then fetuses clearly fall below the threshold. But I have conceded that this account of the threshold is speculative rather than firmly anchored in some larger conception of the nature of

morality. Should this uncertainty leave me agnostic on the question whether foetuses lie within the scope of the morality of respect? Not altogether; for we can say with confidence that, whatever the basis of respect is, it is absent in the case of animals. If, for example, we consider a further variant of the case of the two soldiers in which one soldier wantonly kills a person (by which I mean an individual who passes a certain threshold with respect to a range of psychological attributes, notably self-consciousness, rationality, and so on) while the other kills an animal, it is clearly unacceptable to suppose that the killing of the animal is equally wrong as the killing of the person, no matter what kind of animal it is. But there are only two possibly relevant respects in which a human foetus differs from a non-human animal. One is its membership in the species *homo sapiens*. The other, which derives from its genetic endowment as a member of this species, is the fact that it has various potentials. While it seems obvious that mere species membership is an inadequate basis for respect, potential – particularly potential for rationality and autonomy – may be more a plausible candidate. Rawls, for example, is quite explicit in claiming that the basis of respect (which in his theory is ‘moral personality’) is a potentiality rather than an actual set of capacities.<sup>15</sup>

While it seems to me that potential can indeed be morally relevant, its relevance is exhausted within the morality of interests. One can have an interest in fulfilling one’s potential. I believe, for example, that the foetal self that exists during the later stages of pregnancy has the potential to become a person and that, since becoming a person would be good for it, it also has an interest (however attenuated as a result of diminished psychological connectedness) in fulfilling that potential. The late-term foetus can thus be harmed by being prevented from becoming a person.<sup>16</sup> But its potential alone does not seem a plausible basis for respect. Imagine that we were to discover that dogs in fact have the potential to be self-conscious and rational. No one has realized this until now because no dog has ever realized this potential; for, in order to elicit the potential, it is necessary for someone to cultivate and nurture the relevant capacities through an intensive and highly structured programme of ‘cognitive therapy’. Only through many years of patient work, taking virtually every waking hour of the dog’s life from earliest puppyhood on, can the relevant mechanisms latent in

<sup>15</sup> Rawls, p. 505. Compare Charles Taylor’s observation that, in the Kantian tradition, ‘what is picked out as of worth ... is a *universal human potential*, a capacity that all humans share. This potential, rather than anything that a person may have made of it, is what ensures that each person deserves respect.’ Taylor, ‘The Politics of Recognition’, *Multiculturalism and The Politics of Recognition*, ed. Amy Gutmann, Princeton, 1992, p. 41. Emphasis in the original.

<sup>16</sup> This view is developed in more detail in my ‘The Metaphysics of Brain Death’, *Bioethics*, ix (1995) § V, and more fully still in *Killing at the Margins of Life*.



the dog's brain be developed and activated. For the first five years there are no perceptible results (which is why the process was not discovered earlier), but after ten years, dogs subject to this programme have the cognitive and emotional capacities characteristic of a normal four-year-old human child.<sup>17</sup> If this were in fact discovered to be true, should we conclude that all dogs (or at least all young dogs) deserve respect by virtue of their potential, and indeed that we and our forebears have been guilty of monstrous injustices to dogs, who have always been within the scope of the morality of respect though we have been unaware of it? Clearly not; while respect would be owed to those dogs whose potential was realized, we would not be required by the knowledge that all dogs had this potential to reassess our estimation of the actual worth of all those dogs whose potential was never cultivated. Potential for worth is not the same as worth itself.

There is an analogy here with the more familiar concept of respect as an attitude involving admiration that may be merited to varying degrees, on the basis of achievements or accomplishments, and that may therefore be either earned or forfeited. Here too potential is largely irrelevant to respect. For example, a person merits respect as an athlete only through notable athletic achievements or performances; having the potential to be a great athlete is at best only a weak basis for respect if, for whatever reason, one fails to realize that potential.

We now have a highly conditional argument for the conclusion that fetuses fall below the threshold of respect. If the potential for having a certain level of overall psychological capacity is not a basis for respect, and if fetuses do not differ from animals in any other way that is relevant to respect, then fetuses, like animals, are outside the scope of the morality of respect. This argument is obviously far from decisive, but it does seem to establish a significant presumption.

This argument, of course, has wider implications. Congenitally severely mentally retarded human beings do not even have the potential that normal human fetuses have. It is therefore even less plausible to suppose that they lie above the threshold of respect than it is to suppose that fetuses do. Mildly and even moderately mentally retarded human beings are another matter altogether, though the need to distinguish the severely mentally retarded below the threshold from the moderately retarded above it obviously pressures us to locate the threshold with some precision.

This conditional argument also applies to newborn human infants,

<sup>17</sup> This thought-experiment is obviously reminiscent of Michael Tooley's case involving a kitten injected with a drug that will cause it to develop into a cat with a brain like that of an adult human being. See his 'A Defense of Abortion and Infanticide', *The Problem of Abortion*, ed. Joel Feinberg, Belmont, CA, 1973, pp. 86–8.

implying that they too lie below the threshold of respect. This, of course, is harder to accept than the corresponding conclusion about fetuses. Nevertheless, it is worth noting that our intuitions about the moral status of infants and infanticide are mixed and indeed probably mutually inconsistent. On the one hand, we are moved by our sense of the innocence, defencelessness, and indeed helplessness of infants to believe that they merit special care and protection. Thus if, in yet another variant of the now familiar case, one of the two soldiers were wantonly to kill a normal adult villager while the other were wantonly to kill a newborn baby, many would find the two killings equally wrong and some would even find the killing of the baby to be *more* seriously wrong.<sup>18</sup> On the other hand, however, it has for years been routine practice for doctors to allow certain infants with Down's Syndrome to die when a routine operation could easily have saved them. And, in most cases at least, the motivation for this practice surely cannot have been a desire to spare the child from a life that would not have been worth living. Rather, the reason has been to spare the parents and others the burden of having to care for the child. This practice is widely approved; yet no one would accept a comparable practice of allowing *older* Down's children to die of easily correctable conditions in order to lift the burden from their parents, even assuming (what is most improbable) that the parents would welcome the death. Reflection on this kind of case thus reveals intuitions about infants and infanticide that seem starkly at variance with those that emerge when we consider such practices as killing babies in war.

Since I accept that animals fall outside the scope of the morality of respect and also doubt that the potential for having certain cognitive capacities can serve as a basis for respect, I am disposed to accept that human infants with cognitive capacities comparable to those of non-human animals must also be below the threshold of respect. I am therefore inclined to try to explain intuitions that appeal to the innocence and helplessness of infants as illusions. This strategy does not, however, commit me to the view that infanticide is no worse than the killing of an animal with comparable psychological capacities. For, according to the Two-Tiered Account, the wrongness of killing a being below the threshold is a function of the degree of harm that the killing causes to its victim. And infants, because they lose virtually the whole of a human life, suffer a vastly greater harm in dying than any animal does. And, of course, the side-effects of infanticide are normally incomparably graver than those of the killing of an animal. So it is possible to develop a weighty case against the permissibility of infanticide in

<sup>18</sup> I am indebted to C. A. J. Coady for pressing me with the latter view.

normal cases even if we assume that infants lie below the threshold of respect.

And what about slightly older children? Here it seems obvious that the more orthodox Kantian account sets the threshold too high. If individuals must be sufficiently rational and autonomous to legislate maxims for the governance of their conduct in order to be worthy of respect, then even three- and four-year-olds will fall well below the threshold. Something less than a robust capacity for autonomy must be sufficient.

Focusing on children of different ages suggests an important challenge to the Two-Tiered Account. I will close by exploring this challenge, having obviously left a great many problems unresolved. According to the Two-Tiered Account, all those above the threshold of respect command equal respect. Respect is, in effect, all-or-nothing. Therefore the threshold separating those who command respect from those who do not is not a line but a chasm. Morally, the gap between those above the threshold and those below it is immense. Yet it seems that each of us must cross it with a single small step. This seems to follow from the nature of human psychological development. For there is no moment when one is instantaneously transformed into a rational, autonomous person. Rather, one becomes rational and autonomous by a process that is gradual, continuous, and above all slow. If there is a threshold, no one crosses it with a quantum leap, or any other kind of leap. Yet according to the Two-Tiered Account, there must be some point at which each of us is transformed from a being that merits no respect to one that merits full, maximal respect. This is clearly quite implausible.

There are two ways, each compatible with the other, in which the Two-Tiered Account can be revised in order to mitigate the implausibility. One is to suggest that the threshold takes the form, not of a sharp line, but of a broad band in which it is indeterminate whether an individual merits respect. Thus no child goes to bed one night as a being unworthy of respect only to wake up the next morning above the threshold. Rather, there is a period in early human development during which the relevant cognitive and emotional capacities are developing and in which it is neither true nor false that the individual commands respect. With this revision, the Two-Tiered Account no longer implies that there must be a moment of transformation. But a problem remains. For, even if the band of indeterminacy spans a period of months, there will be a wide divergence between the moral status of the child before this period and after it. For the child still goes from being unworthy of respect to commanding the highest respect without passing through any intermediate state. There is, in short, an enormous discontinuity in the child's moral status, even if it is not

temporally abrupt; but this discontinuity is not paralleled by a correspondingly significant discontinuity in the child's psychological development or maturation.

Since the Two-Tiered Account cannot do without a threshold of respect, it unavoidably implies that each of us undergoes a leap in moral status sometime early in life. But the leap can be narrowed if we return to the idea, rejected earlier, that there can be degrees of worth above the threshold. If we assume that an individual's worth continues to vary with his or her cognitive and emotional capacities above the threshold, then we can say that people may command varying levels of respect. If we concede this, then the moral transformation that occurs when a small child crosses the threshold of respect is minimized. Instead of making a quantum leap from being altogether unworthy of respect to being worthy of the highest respect, the child goes from commanding no respect to commanding only minimal respect and from there on to ever-increasing levels of respect until his or her cognitive and emotional capacities are fully realized.

This suggested revision is supported by other considerations. There is, for example, a disturbing arbitrariness in the idea that a being's worth varies with the degree to which its cognitive and emotional capacities are developed if the being is below the threshold but not if it is above it. Moreover, many of the relevant cognitive and emotional capacities are actually possessed by animals though to a much lower degree. But, if what separates us from animals is really only that we possess certain capacities *to a higher degree*, then it seems all the more arbitrary to insist that further differences of this same sort cannot distinguish some of us from others.

This revision is not, however, without cost. As I noted earlier, if we accept that worthiness of respect varies with a person's cognitive and emotional capacities, then, unless we shift to an altogether different concept of respect, we will be committed to the shockingly and invidiously élitist view that it is more seriously wrong, other things being equal, to kill people with higher cognitive and emotional capacities than it is to kill those whose capacities are lower. We therefore appear to face a profound dilemma. If, on the one hand, we accept that worth varies with capacities even above the threshold of respect, the Two-Tiered Account can then offer a more plausible account of the moral status of young children; but it will also have objectionably inegalitarian implications for the morality of killing adults. If, on the other hand, we insist that worth is constant above the threshold, the Two-Tiered Account will then successfully capture important intuitions about the equal wrongness of certain killings; yet it will not only treat variations in cognitive capacity above and below the threshold in arbitrarily different ways but also be unable to offer an account of the moral

status of children that maps comfortably onto the facts of human psychological development.

How we should respond to this dilemma is yet another problem that this paper must leave unresolved.<sup>19</sup>

<sup>19</sup> An earlier version of this paper was read at the Australian National University. I am grateful for the audience's comments on that occasion. I have also greatly benefited from comments by Roger Crisp and Brad Hooker.