

JEFF MCMAHAN

CHALLENGES TO HUMAN EQUALITY

(Received 9 July 2006; accepted in revised form 9 July 2006)

ABSTRACT. According to liberal egalitarian morality, all human beings are one another's moral equals. Nonhuman animals, by contrast, are not considered to be our moral equals. This essay considers two challenges to the liberal egalitarian view. One is the "separation problem," which is the challenge to identify a morally significant intrinsic difference between *all* human beings and *all* nonhuman animals. The other is the "equality problem," which is to explain how all human beings can be morally equal when there are some human beings whose psychological capacities (and, in some cases, their psychological potentials as well) are no higher than those of certain nonhuman animals. The focus throughout is on the ethics of killing but the arguments are of broader relevance. The essay reaches a skeptical conclusion about our ability to meet these challenges.

KEY WORDS: animals, cognitive disability, equality, killing

1. THE SEPARATION PROBLEM

In contemporary Western societies, common sense morality is liberal egalitarian in character. With the possible exception of human beings in their embryonic or fetal stages, all human beings are recognized – in principle if not in practice – as one another's moral equals. Each human being matters equally; each has equal value and equal human rights. One element of this liberal egalitarian view is the belief that all wrongful killings of human beings are *equally* wrong, except, perhaps, when there are relevant differences in the state of the agent, such as the difference between a killing that the agent intends and one that the agent merely foresees as a side effect of his action.¹ In the book

¹ The object of evaluation here is the causing of a human being to cease to exist, not the manner in which the human being is killed. It may well be more seriously wrong to kill a person in an agonizing manner than to kill the same person painlessly. But this is because the former involves two distinct wrongs: causing the agony and causing the death. The claim of equal wrongness applies to the causing of death alone.

ably and generously reviewed by Rahul Kumar in the preceding article, I discussed a more restricted view, which I called the “equal wrongness thesis.”² The equal wrongness thesis applies not to all wrongful killings of human beings but only to wrongful killings of *persons* – that is, individuals with psychological capacities beyond a certain threshold of self-consciousness and minimal rationality. According to the equal wrongness thesis, the extent to which it is wrong to kill a person is unaffected by facts about the victim other than those that may make him or her *liable* to be killed, or *deserve* to be killed, if indeed it is possible for a person to deserve to be killed – that is, facts other than those that may make the victim non-innocent in the relevant sense. It is, according to the equal wrongness thesis, no less seriously wrong to kill an old person than to kill a young person, or to kill a person of melancholy disposition than to kill a person with a happy temperament. The extent to which it is wrong to kill a person is, in short, unaffected by the degree of loss or harm he would suffer by being killed. Provided that he satisfies the criteria of personhood, the wrongness of killing him is also unaffected by his nature or capacities – for example, by whether he is virtuous or vicious, likable or obnoxious, sensitive or insensitive, intelligent or unintelligent, and so on. In the book, I neither endorsed nor challenged the equal wrongness thesis, though my discussion of it was respectful and I assumed it to be true for the sake of argument.

Most liberals insist that decisions not only about killing but even about saving lives must not be based on “quality of life” considerations, or “quantity of life” considerations. There is, however, some disagreement about this, even among liberals, in cases other than those involving *wrongful* killing. Some accept that when one can save some but not all the people who will otherwise die, it can be permissible, at least in some instances, to give priority to those who would lose most by dying. And some accept that this can be permissible as well in choices of whom to kill when circumstances make it permissible to kill someone. Suppose, for example, that a runaway train is on a track leading to the station where another train full of passengers has stopped. If nothing is done, the runaway train will crash into the stationary train, killing hundreds of people. But the train can be diverted onto one of two branching tracks before it reaches the station. On one of these tracks there is a 20-year-old who

² Jeff McMahan, *The Ethics of Killing: Problems at the Margins of Life* (Oxford: Oxford University Press, 2002), pp. 233–248.

cannot get off, while on the other there is a 90-year-old who cannot get off. Most of us believe that it would be permissible to divert the train onto one of the two tracks. Some liberals accept that it would be permissible, or perhaps even required, to choose to kill the older person on the ground that she would lose less by being killed, if there were no weightier considerations that favored killing the younger person instead. But many liberals believe that even in this kind of case, equality requires that each of the two potential victims be given an equal chance of surviving, so that, ideally, a randomizing device, such as flipping a coin, should be used to determine onto which track the train should be diverted. These people believe that a killing that would be permissible if the choice of victim were made randomly would be wrongful if the choice were made on the basis of a comparison between the potential victims' characteristics.

The equal wrongness thesis, which applies to wrongful killings of persons, is deeply intuitive. Yet it is too weak to capture most people's intuitions, which extend the scope of equality beyond persons to include all human beings, or at least all postnatal human beings. Almost no one, however, believes that nonhuman animals are our moral equals or have rights equal to ours. They are by definition ineligible for human rights. And virtually everyone agrees that killing an animal cannot be as seriously wrong as killing an innocent human being, or a person. Not only are animals not our equals, but they are not even equal among themselves; thus it is regarded as more seriously wrong to kill a higher nonhuman animal, such as a chimpanzee, than it is to kill a lower animal, such as a lizard.

When pressed to explain why animals lie outside the scope of liberal egalitarian principles, such as the equal wrongness thesis, that are thought to govern our treatment of other human beings, most liberals respond initially by appealing to certain psychological capacities that human beings possess but that animals lack. Human beings, they point out, have capacities for self-consciousness, rationality, autonomy, the use of language, action on the basis of reasons, and so on. Some one or combination of these capacities is what relevantly distinguishes human beings from animals and provides the foundation for human equality. Yet according to the common understanding of what it is to have a capacity, some human beings lack all such capacities. Some, such as the severely demented and the irreversibly comatose, lack them now but had them in the past. Others, such as most fetuses and newborn infants, lack them

now but have the potential to have them in the future.³ But some human beings – namely, those that are congenitally and radically cognitively impaired – have never had the capacities and also seem to lack the potential to have them. And even among those human beings who possess some of the capacities, there are some who do not possess them all, while among those who possess them all, there are some in whom the relevant capacities are more highly developed than they are in others. How, then, can such capacities provide a basis for human equality?

There are here two related but distinguishable problems for liberal egalitarianism. One is to defend the common sense view that all human beings are owed a form of consideration that is different from and higher than that which is owed to other animals. The other is to show how that form of consideration can be owed equally to all human beings. I will call these the “separation problem” and the “equality problem,” respectively. I believe that the separation problem cannot be solved. If some individuals are owed a higher form of consideration than is owed to any animal, that must be because they have certain higher psychological capacities. An acceptable form of separation may place most human beings above a relevant threshold of psychological capacity, but it will leave some human beings below it and may place some nonhuman animals above it. And if the threshold is defined by psychological capacity, the equality problem will remain, though not with respect to all human beings but with respect to all those individuals above the relevant threshold.

Most philosophers believe that both these problems for liberal egalitarianism can be solved. They believe that we can identify a basis for distinguishing morally between human beings and nonhuman animals that is not variable. They accept the relevance of psychological capacities but claim that, while possession of higher psychological capacities is sufficient for inclusion within the scope of liberal egalitarian principles, such as the equal wrongness thesis, it is not necessary. Human beings who lack the relevant capacities may nevertheless be within the scope of the principles by virtue of having a closely related characteristic. This type of response is defended by Kumar, who contends that “claims about the life-cycle of a particular

³ Liberal egalitarians who believe that abortion can be permissible may take the fact that fetuses lack the relevant capacities as partial confirmation that it is these capacities that distinguish most human beings from other animals. But if so, they have a problem if they share the common opposition to infanticide (See Jeff McMahan, “Infanticide,” *Utilitas*, forthcoming).

kind of living thing, or species, are just *constitutive of what it is to be a member of that species,*” and that liberal egalitarian principles are “responsive, not just to the particular properties of an individual, but to the characteristic life-cycle of the species to which that individual belongs.”⁴

There are two ways in which Kumar’s remarks can be interpreted. In one passage he writes that

claims about species are not statistical generalizations. Rather, what they concern is the essential nature of a living kind, revealing facts about the *normal life-cycle* of that kind of living thing. The use of “normal” here is unashamedly normative. ... What respect for the value of a living thing requires will depend on the characteristic life-cycle, or nature, of members of that [individual’s] species.⁵

This passage suggests that in order to have the same moral status as a normal human being, it is not necessary for an individual actually to have a rational nature or to be internally directed toward the development of a rational nature. Instead, these properties are *normatively* characteristic of human beings – that is, all human beings *ought* to have them even if there are some that do not. Because of this, those human beings who will not have the normal or characteristic human life-cycle, or who do not have the nature characteristic of human beings, must nevertheless be accorded the same respect that would be due to them if they did.

This view raises the question why facts about the nature of some individuals could determine how other individuals that lack that nature ought to be treated. Somehow membership in the same biological kind is supposed to produce the requisite moral alchemy, but it is opaque to me how this is supposed to work. I will not, however, pursue this issue further here, as I have done so elsewhere.⁶ I will also not address certain other questions raised by this interpretation, such as whether there is any basis for what is normatively characteristic of a species other than what is statistically characteristic of it, and, if not, why we should suppose that what is statistically characteristic is invested with normative significance.

I will focus instead on a second interpretation of Kumar’s view. According to this interpretation, what Kumar claims is that all human beings are equal in their possession of the essential nature of human

⁴ Rahul Kumar, “Permissible Killing and the Irrelevance of Being Human,” *The Journal of Ethics*, this issue (cited from two different paragraphs). Emphasis added.

⁵ Kumar, “Permissible Killing and the Irrelevance of Being Human.”

⁶ Jeff McMahan, “‘Our Fellow Creatures,’” *The Journal of Ethics* 9 (2005), pp. 355–359.

beings – the nature that defines their characteristic life-cycle and includes having (or, perhaps, having had) a capacity for “rational self-governance,” or *being internally directed toward the development of such a capacity*.⁷ It is the possession of this essential nature that brings all human beings within the scope of liberal egalitarian principles.

This view has much to recommend it. If successful, it brings all human beings within the scope of liberal egalitarian principles, while excluding all or most nonhuman animals. This does not mean, however, that the view is committed to endorsing common sense morality’s assignment of a vastly lower status to animals. It could deny that most animals are our moral equals and yet concede that perhaps some few of the higher animals are, while claiming that others are worthy of significantly better treatment than they are accorded by current practices. Admittedly, the view would disqualify liberals who defend the permissibility of abortion from appealing for support to claims about fetal status, such as the claim that fetuses are not persons; yet these liberals could, as Kumar suggests, appeal to arguments that focus instead on the fetus’s lack of a right to the use of the pregnant woman’s body as a means of life support. Unlike arguments for abortion that appeal to fetal status, this sort of argument has no tendency to support the permissibility of infanticide – surely an advantage if we want to preserve deeply held intuitions.

Many of Kumar’s claims refer at least implicitly to biologically immature human beings that lack the present capacity for rational self-governance but nevertheless have the potential to develop that capacity. He claims that respecting these individuals’ nature as human beings involves not impeding and even facilitating the full realization of their potential and thus their essential nature. In the case of a human being that in a clear sense lacks even the potential for rational self-governance, such as a fetus with defective genes that direct the growth of the brain, he believes that respecting its nature involves providing that potential if possible – for example, by supplying it with properly functional genes for brain growth. For given its nature as a human being, its lack of those genes is a

⁷ Kumar, “Permissible Killing and the Irrelevance of Being Human.” For a similar view, see Robert P. George and Alfonso Gómez-Lobo, “Statement of Professor George (Joined by Dr. Gómez-Lobo),” an appendix to *Human Cloning and Human Dignity: An Ethical Inquiry*, a report by the President’s Council on Bioethics, available at http://www.bioethics.gov/topics/cloning_index.html (accessed on 30 June 2006). And compare S. Matthew Liao, “Virtually All Human Beings as Rightholders,” unpublished manuscript.

misfortune, though the absence of those genes, and the associated apparent lack of potential for rational self-governance, are not misfortunes for a nonhuman animal.

There is a challenge here that I have not adequately addressed either in my book or in a recent article in *The Journal of Ethics* in which I advanced various arguments against views, such as Kumar's, that claim that "the morality of our treatment of an individual ought to be... guided by the moral norms that regulate how members of that individual's species ought to be treated."⁸ The challenge derives from the view that the reason why all human beings are included within the scope of liberal egalitarian principles, such as the equal wrongness thesis, is that they all have, as part of their essential nature or constitution at all times at which they exist, either certain higher psychological capacities, such as the capacity for rational self-governance, or an inner directedness toward the possession or development of those capacities. On this view, membership in the human species is not morally significant only, or primarily, or perhaps at all, because it involves a significant relation to other human beings. It is significant because it is sufficient for the possession of an intrinsically significant nature, whether or not that nature is fully realized.⁹

Most mature human beings have a rational nature and most immature human beings do seem to be inherently or internally directed toward the development of the capacity for rationality. Suppose we grant – what I believe to be plausible – that the actual

⁸ Kumar, "Permissible Killing and the Irrelevance of Being Human." See McMahan, "Our Fellow Creatures'."

⁹ What about human beings who once had higher psychological capacities but have irreversibly lost them? There are two ways in which this view can attempt to accommodate such individuals within the scope of liberal egalitarian principles. One is to claim that even if the *acquisition* of a certain moral status requires the possession of or potential for certain capacities, that status can nevertheless survive the loss of those capacities and even the loss of the potential for those capacities. The other is to claim that even if the areas of the brain that are the physical basis of an individual's higher psychological capacities are destroyed, that individual remains genetically or otherwise constituted to be directed toward the *possession* (though not the *development*) of those capacities as long as he or she remains alive. I think the former of these two claims is more plausible, in part because it provides a basis for the possibility of wronging the dead. Moral claims can not only survive an individual's loss of capacities the possession of which was once necessary for the individual to be able to exert those claims, but can even survive the death of the person who is the source of the claims.

possession of a rational nature is sufficient to bring an individual within the scope of liberal egalitarian principles.¹⁰ There remains the familiar question about the potential for rationality as a basis for this moral status: namely, why should an immature human being's internal directedness toward the development of a rational nature affect how we ought to treat that individual now? There is, of course, a good answer to that question in cases in which how we treat the individual now will affect her for better or worse later, when she actually has a rational nature. But there are certain types of act the effects of which are limited entirely to the present, including – crucially – acts of killing. Why should the morality of an act of killing be governed by the kind of respect that is appropriate for a nature that the individual killed does not have now but may have later, though *only* if it is not killed? Why should an act of killing not be governed instead by due consideration for the nature of the individual at the time of action (or of the death, if it occurs later)?

I believe, as I argued in *The Ethics of Killing*, that there are no good answers to these questions. I will not rehearse those arguments here but will instead raise what I think is a more difficult question. What reason is there to suppose that all human beings are in fact internally directed or programmed toward the development of a rational nature? There are some human beings – those who are congenitally and radically cognitively impaired (henceforth the “radically impaired”) – who in at least one obvious sense lack the potential for the development of a rational nature. While most immature human beings (embryos, fetuses, newborn infants) will, given a favorable environment, develop a capacity for rationality, those immature human beings that are radically cognitively impaired cannot develop this capacity, even with the most extensive forms of assistance that we are currently able to provide.

For Kumar's suggested strategy to succeed, there must be a clear and morally significant sense in which the present nature of the radically impaired is internally directed toward the development or realization of a capacity for rationality, or rational self-governance. Perhaps the most plausible claim for the defender of this strategy to make is that the inherent tendency toward rationality is present in the human genome, which the radically impaired share with the rest of

¹⁰ Perhaps there are exceptions in rare cases, such as extreme instances of psychopathy, in which a human being has a developed capacity for instrumental rationality but utterly lacks certain other capacities necessary for moral agency. I will not explore this possibility here.

us. Their essential nature is thus inherent in their genes, but its development or expression is impeded because the action of the genes that direct the proper growth of the brain is somehow blocked. It may be that their action is countermanded by some other malfunctioning genes, or that something that is required for their activation is missing, or that some of them are defective in some way.

If the genes for the proper growth of the brain are present but something prevents their activation or blocks their action, there is a clear sense in which the potential for a rational nature is present. But it is less clear how long that potential persists. For the potential consists in the ability of the genes to direct the growth of a normal human brain, which is the physical basis of the capacity for rationality. But if the action of the genes is impaired and the brain grows abnormally, it appears that the potential for the growth of a normally functioning brain has been lost. For the genes cannot direct the growth of the brain once it already exists. If, even at this point, those genes could direct the reconstruction or augmentation of the defective brain in an identity-preserving way, then the potential, or inner directedness, toward a rational nature would remain; but to the best of my knowledge there is no reason to believe that the relevant genes have that capacity. If that is right, then radically impaired human beings whose condition results from the blocked functioning of normal genes lose their internal directedness toward the possession of a rational nature very early in life, when the development of their brains passes a certain point.

What if the defective growth of the brain results not from the blocked action of normal genes but from defects in the genes themselves? Is it nonetheless the nature of these radically impaired human beings to be rationally self-governing? It seems as if these individuals were never internally directed toward the development of the capacity for rationality but were always directed toward the development of the cognitive capacities that they in fact have. Consider an instance in which defective human genes direct a human brain to grow in such a way that, when fully developed, its cognitive powers are no higher than those of some lower animal. Certainly in functional terms, and perhaps in structural terms as well, these genes more closely resemble the corresponding genes of a lower animal than they do their normally functioning human counterparts. Indeed, because these defective human genes function to produce a brain with powers of cognition similar to those of a lower animal, they are *less* similar in functional terms to their properly functioning human

counterparts than are the corresponding genes in a higher animal, which function to produce a brain with cognitive capacities closer to those of a normal human brain. On what basis, then, can it be claimed that it is the inherent nature of the individual whose genes these are, but not the nature of a higher animal, to be internally directed toward the realization of a capacity for rationality?

We have thus far considered that radical impairment could result either from the suppression of the action of certain genes that carry the code for the development of the human brain, or from defects in the genes themselves. But there is another possibility. Some of the relevant genes could be entirely absent. Whether such cases occur or have occurred naturally is a question on which I have no information. But even if they do not occur spontaneously in nature, they could presumably be created by manipulating a pair of human gametes prior to fertilization in a way that would ensure the deletion of certain genes necessary for directing the normal growth of the brain. The resulting radically impaired individual would certainly be a human being by any reasonable criterion of species membership, whether that criterion referred to genes, genealogy, capacity for interbreeding, or whatever. There would then be a human being without any genetic basis for the development of a rational nature. It would therefore be hard to make sense of the idea that this human being would be internally directed toward the full realization of its inherent nature as a rational being.

Suppose there were a human fetus in this condition – that is, one lacking some of the genes necessary for the development of a brain with the capacity for rational thought. There is one sense in which this individual might have the potential for a rational nature. Suppose there were a therapy that could supply the missing gene or genes and that if it were administered to the fetus, that fetus would grow to produce a person with normal cognitive capacities rather than being radically impaired. And suppose further that the transformation would be identity-preserving, in that the person would be the same individual that the radically impaired individual would have been. The therapy, in other words, would determine whether one and the same individual would have normal or radically impaired cognitive capacities. The claim that this genetically defective fetus has the potential for a rational nature is therefore just the claim that this very same individual could become rational. But even though this human fetus could become a rational being, it would not be *internally* directed toward the realization of a rational nature.

Using the terms I proposed in the book, we could say that its potential for a rational nature was extrinsic rather than intrinsic. For the physical basis for the potential would have to be imported. The aspect of the potential that would be inherent would be only a certain receptivity to an externally generated but identity-preserving transformation into a being with a rational nature.

Many animals, it seems, also have the same kind of potential to have a rational nature. This is true if there is in principle a genetic therapy that could transform them into rational beings. It is true even though no such therapy now exists. All that is necessary for an animal to have this potential is that such a therapy is physically possible. This is clearly not the kind of potential that Kumar and others believe can be a sufficient basis for inclusion within the scope of liberal egalitarian principles, such as the equal wrongness thesis. Yet it seems that this is the only kind of potential for a rational nature that a human individual could have if that individual lacked some of the genes necessary for the development of the brain beyond that of a radically impaired human being. And it may be the only kind of potential possessed by a human being with defective versions of certain genes necessary for the development of such a brain. Neither type of human being seems internally directed toward the realization of a rational nature.

One option is to hold fast to the claim that the basis for inclusion within the scope of liberal egalitarian principles is either the possession of a rational nature or an internal directedness toward developing such a nature. If I am right that individuals lacking genes for the growth of a brain that is not radically impaired, and possibly also those with genes that because of defects cannot produce a brain that is not radically impaired, are not internally directed toward a rational nature, then we must either concede that some human beings lie outside the scope of liberal egalitarian principles or else deny that these individuals are human beings. I assume that the latter is not a genuine option, since no one supposes that membership in the human species is determined by the presence or absence of normally functional genes for the growth of the brain.

Suppose one were to accept that only those human beings with normally functioning genes for the growth of the brain are within the scope of liberal egalitarian principles. There would still be problems. One is that it is hard to believe that a difference at the level of genes could have such significance. On this view, a radically impaired

human being whose impairment has resulted from the blockage of the action of normal genes would be our moral equal while another radically impaired human being whose impairment was the result of a missing or defective genes would not be, *even* if the latter's actual cognitive capacities were higher.¹¹

A second problem is that internal directedness toward a rational nature is not all-or-nothing, but is instead a matter of degree. In a paradigmatic case of internal directedness or intrinsic potential, all that is needed from an external source for the realization of the potential is nutrition, hydration, shelter, and so on. But there is then a spectrum of possible cases in which in each succeeding case just a little more is needed from the outside for an individual to develop a rational nature. There is no threshold that marks a sharp separation between cases of intrinsic potential and extrinsic potential. This is a problem if only those human beings with the intrinsic potential for a rational nature can be within the scope of liberal egalitarian principles. For the application of the principles now rests on a distinction that cannot be drawn with precision.

I argued in *The Ethics of Killing* that intrinsic potential for a rational nature, which I assume to be equivalent to internal directedness toward the development of a rational nature, is not a

¹¹ In a seminal paper on abortion, Michael Tooley advanced a complicated argument against the moral significance of potential based on a thought-experiment involving a chemical that, if injected into a kitten's brain, would give that kitten the potential to develop into a rational being [Michael Tooley, "A Defense of Abortion and Infanticide," in Joel Feinberg, (ed.), *The Problem of Abortion* (Belmont: Wadsworth Publishing Company, 1973), pp. 86–88]. This thought-experiment has force against the moral significance of the distinction between intrinsic and extrinsic potential. Updating the example for advances in genetics, we can imagine a gene therapy that, if administered to a kitten, would give that kitten the genetic basis for the development of a rational nature in an identity-preserving way. Prior to the administration of the therapy, the kitten would have only extrinsic potential for a rational nature. But after the therapy had been administered, its potential would be intrinsic: it would then be internally directed toward the development of a rational nature. Suppose that immediately after the administration of the therapy, it would be possible to reverse its effects, causing the kitten to revert to an ordinary kitten with only the extrinsic potential for a rational nature. I think there would be no more reason not to reverse the effects of the therapy than there would be to administer it in the first place. And I agree with Tooley that there would be no more reason not to kill the kitten immediately after it had received the therapy than there would be before. But these intuitions, which I believe would be widely shared, would be indefensible if intrinsic potential were morally more significant than extrinsic potential.

plausible basis for the sort of worth that brings an individual within the scope of liberal egalitarian principles, such as the equal wrongness thesis. I have argued here that even if it were a sufficient basis for that sort of worth, not all human beings have it. So what I claimed in the book still seems to me to be true: namely, that whatever it is that makes us members of the human species is not in itself morally significant, though it coincides in most cases with properties that are highly significant.

2. THE EQUALITY PROBLEM

Let me restate our original problem. We are committed to certain principles of equality, such as the equal wrongness thesis. But animals are excluded from the scope of these principles; we do not think that they are our moral equals or that to kill an animal is as seriously wrong as it is to kill a human being, or a person. It is plausible to think that what distinguishes us morally from animals is that we have certain psychological capacities that they lack, or that they possess in only a very rudimentary form. Yet there are some human beings that most of us believe to be in an important sense our moral equals that also lack these capacities, or possess them only in the primitive forms in which they are present in some animals. The only way to ensure that all human beings are within the scope of our principles of equality while excluding most or all animals is to claim that whatever it is that makes us members of the human species is also what makes us one another's moral equals, or at least is invariably correlated with what makes us one another's moral equals. But I have argued that neither of these claims is true.

Part of the problem is that the relevant principles of equality are all-or-nothing. Either an individual is or is not our moral equal. The satirical force of the slogan in G. Orwell's *Animal Farm* – “Some Animals Are More Equal Than Others” – depends on our recognition that there are no degrees of equality. If I am right that there is nothing in or invariably correlated with membership in the human species that can be the basis of our moral equality, then whatever the criterion of equal moral status is – and particularly if it is connected with psychological capacity – either some human beings are not our moral equals or some animals are.

In *The Ethics of Killing* I made an effort to preserve our commitment to equality – specifically, the equal wrongness thesis – within these

constraints. I proposed a “two-tiered” account of the morality of killing, according to which the killing of individuals below a certain threshold of psychological capacity – which I called the “threshold of respect” – is governed solely by a proper concern for their good (technically, their time-relative interests – a notion that is elucidated in Kumar’s review), whereas the killing of individuals above the threshold is governed by a requirement of respect for their inherent worth. I suggested that while the degree to which it is wrong to kill individuals below the threshold varies with the strength of their time-relative interest in continuing to live, the wrongness of killing individuals above the threshold does not vary with the strength of their interests or with any features of their nature. For individuals above the threshold have *equal* worth, so that wrongful killings of these individuals are always equally wrong, provided that the individuals are innocent in the relevant sense and facts about agency are held constant.

This approach, which attempts to preserve the equal wrongness thesis, has been powerfully criticized in a critical notice of *The Ethics of Killing* by Tim Mulgan. While he advances various specific objections, the larger general objection is that the all-or-nothing moral claims that I attempt to defend, such as the equal wrongness thesis, cannot map cleanly onto the characteristics that are assumed to be their basis and on which they are supposed to supervene, since those characteristics come in degrees.¹² Thus, he writes that “McMahan wants to combine a naturalistic, broadly Humean, picture of a world where continuous properties come in degrees, with a set of Kantian intuitions that clearly require sharp boundaries between persons and non-persons. This is an essentially unstable combination.”¹³ In this he is obviously right. This is, in fact, an

¹² See Tim Mulgan, “Critical Notice of *The Ethics of Killing*,” *Canadian Journal of Philosophy* 34 (2004), pp. 443–460. It is perhaps worth noting here a couple of mistaken assumptions that Mulgan makes about the nature of my position. He assumes that, like Peter Singer, I attribute only impersonal value to the satisfaction of the interests of beings that lack the capacity for self-consciousness, so that these beings are, in effect, “replaceable” (Mulgan, “Critical Notice of *The Ethics of Killing*,” p. 455). But I deny that the frustration of one individual’s interests through killing can be made up for by the creation and satisfaction of comparable interests through the creation of another individual. He also assumes that I would regard a 10-year-old child as falling beneath the threshold of respect and thus as being outside the scope of liberal egalitarian principles, such as the equal wrongness thesis. But I accept that any child capable of understanding that it has a future and of acting for reasons is a person and comes within the scope of liberal egalitarian principles.

¹³ Mulgan, “Critical Notice of *The Ethics of Killing*,” p. 458.

instance of a problem that has always plagued theories of human equality – namely that we are held to be normatively equal and our moral status is held to supervene upon facts about our nature, yet there are really no relevant respects in which we are by nature equal.

My attempt to defend the equal wrongness thesis by devising a two-tiered account of the morality of killing is vulnerable to a variety of challenges, some of which I noted in my book. Most derive from the way that the account attempts to base an all-or-nothing moral framework on characteristics that are matters of degree. Here are seven such challenges, all of which I consider to be quite powerful.

- (1) Suppose that there is a threshold of respect and that what distinguishes those who are above it from those below it is that those above it possess certain psychological capacities not possessed by, or not developed to a sufficient degree in, those below it. How can it be that variations in the degree to which those capacities are developed in those above the threshold do not matter at all to the degree to which it is wrong to kill them? If the possession of these capacities, or their possession above a certain level, is what is necessary for inclusion within the scope of the equal wrongness thesis, how could it be that the degree to which these capacities are developed above the threshold is entirely irrelevant?
- (2) Each of us began life below the threshold of psychological capacity that we had to reach in order to become the moral equal of others. We are now above it. But as I have noted, equality is all-or-nothing, and it is implausible to suppose that there was a point when each of us was suddenly transformed into the moral equal of others. For the development and maturation of our psychological capacities is gradual and continuous, without abrupt discontinuities. Given that the difference between being and not being within the scope of liberal egalitarian principles is profoundly significant, it is impossible to make the transition from one type of status to the other through a merely incremental increase in psychological capacity. Because the relevant capacities are matters of degree and lack sharp boundaries, the difference between an individual just above the threshold and an individual just below it cannot be the difference between the possession of some capacity and the failure to possess that capacity.

- (3) Assuming that problem 2 can be overcome and we can identify a sharp threshold dividing those who are our moral equals from those who are not, where does the threshold lie? If it is sufficiently low that very young children are above it, then some nonhuman animals, such as chimpanzees, should be above it as well. If it is high enough to exclude all animals, it will also exclude young children and adults whose psychological capacities are permanently arrested at or below the level of young children.
- (4) Common sense morality accepts that the degree of harm caused to the victim by wrongful acts of nonlethal injuring or harming is relevant to the degree to which the act is wrong. It is, for example, more seriously wrong to break a person's leg than to break a person's toe, or to cause a person to suffer severe pain for a day than to cause a person to suffer the same level of pain for a minute. But the degree of harm caused to the victim by an act of killing may also vary significantly. Normally, for example, a 20-year-old would be harmed to a much greater degree by being killed than a 90-year-old would, because the 20-year-old would suffer a greater loss of valuable life. But the equal wrongness thesis treats this difference in the degree of harm as insignificant. Yet it seems arbitrary to suppose that variations in the degree of harm caused by killing are morally insignificant when variations in degree among other harms are highly significant.
- (5) One specific instance of this more general challenge concerns the harm caused by rendering a person unconscious when the conscious life he would otherwise have would be worth living. It is clearly more seriously wrong to render a person unconscious for a longer period than for a shorter period – for example, for a year rather than for a day. Yet, at least in cases in which the body ages at the normal rate during periods of unconsciousness, the form of loss or deprivation involved in complete unconsciousness is the same, or virtually the same, as that involved in death – namely, the loss of the goods of conscious life. So it is hard to see how differences in the degree of loss a person suffers by being rendered unconscious can affect the degree to which the act is wrong, while the degree of the same kind of loss a person suffers by being killed cannot.¹⁴

¹⁴ I owe this specific challenge to Kasper Lippert-Rasmussen, who develops it with great ingenuity in “Why Killing Some People is More Seriously Wrong than Killing Others”, *Ethics* (forthcoming) .

- (6) There are some cases, involving negligible or nearly negligible loss, in which the equal wrongness thesis has implications that conflict with most people's intuitions. If a person can live only a very short time in any case – for example, if he is certain to die from causes that cannot be prevented in one minute, or an hour, or perhaps even a day – it does not seem that it would be as seriously wrong to kill him – for example, as a means of saving another innocent person's life – as it would be if he had a much longer period of life in prospect.
- (7) As I noted early in this article, many of the sorts of difference among people that the equal wrongness thesis claims are irrelevant to the wrongness of killing nevertheless seem to be relevant in certain related types of choice – for example, choices of whom to save, or whom to kill in cases in which killing is permissible. Recall, for example, the example of the runaway train. To many people, though not all, it seems permissible to divert the train on the track on which the 90-year-old is trapped precisely because he would suffer a lesser loss in being killed than the 20-year-old on the other track would. But if a difference in the degree of harm to the victim can make a moral difference in this type of case, why should it not also make a difference in cases involving *wrongful* killing?

There are of course possible responses to some these challenges. In the case of objection 6, for example, it may be that cases in which the harm from death would be negligible and the killing would serve some important purpose, the killing would not be wrong at all. If so, these cases would not challenge the equal wrongness thesis, which applies only to killings that are wrongful. But this response only increases the force of objection 7, which questions why the degree of harm caused can be relevant to a choice among possible killings when some act of killing would be permissible, but not to the degree to which a wrongful killing is wrong.

Some of these challenges could be addressed by refining the account, making it effectively a three-tiered rather than a two-tiered account. According to this refined account, there are two thresholds along the spectrum of levels of psychological capacity. Individuals below the lower threshold are conscious but uncontroversially lacking higher psychological capacities such as self-consciousness, rationality, and autonomy. Individuals above the second threshold are uncontroversially self-conscious, rational, and autonomous.

Individuals between the two thresholds are at various stages in the gradual development of the higher psychological capacities, or are permanently arrested at one of these intermediate levels of psychological development.

The treatment of individuals below the lower threshold of psychological capacity is governed by a proper concern for their good, or time-relative interests. The interests of these individuals may be traded against one another without any weighting other than for the strength of the interest itself. Their interests, in short, may be treated in a consequentialist manner. By contrast, individuals above the higher threshold have what might be called "maximum inviolability." They are all one another's moral equals and their treatment is governed by constraints that, while not absolute, are as strong as any constraints can be in their application to persons with psychological capacities within the range known to us (It is possible that aliens with psychological capacities substantially higher than those of any human being might have a higher degree of inviolability, so that their treatment would be governed by constraints that were even stronger, though still not absolute). That individuals above the higher threshold have maximum inviolability means, among other things, that if they are innocent in the relevant sense, they may not be intentionally killed or sacrificed for the sake of others except when sacrificing them is necessary to avert some vastly greater harm to other innocent individuals who are also above this higher threshold.

According to the two-tiered account, there are only these two forms of status: violability for the greater good below the threshold of respect and maximum inviolability above it. There is no intermediate status. But according to the three-tiered account, there is. Individuals whose psychological capacities lie between the lower and higher thresholds have an intermediate moral status.¹⁵ They are neither wholly sacrificeable in the service of the greater good nor maximally inviolable. The treatment of these individuals is governed by constraints, but the strength of these constraints varies with the level of psychological capacity of the individual to whom they apply. The treatment of an individual with psychological capacities just barely above the lower threshold would be governed by very weak constraints. Such an individual could permissibly be sacrificed in

¹⁵ In McMahan, *The Ethics of Killing*, pp. 264–265, I proposed a similar but cruder view. The main difference is that the view I suggested in the book held that all individuals with intermediate moral status have the *same* status.

order to prevent a greater harm to others, though the harm to be prevented would have to be slightly more than would be required to justify the sacrifice on consequentialist grounds. Another individual with somewhat higher psychological capacities would be protected by somewhat stronger constraints: in order for the sacrifice of such an individual to be permissible, the harm that would thereby be prevented would have to be somewhat greater. And what it would take to justify the sacrifice of individuals with higher capacities continues to increase with the level of psychological capacity until the point just below the higher threshold is reached. Individuals just below that threshold would have almost maximum inviolability.

This three-tiered account seeks to reconcile our egalitarian intuitions to the greatest degree possible with the fact that the characteristics on which moral status seems to be based are present to widely varying degrees among human beings and animals. It preserves moral constraints above the lower threshold and equality above the higher threshold (so that it is only individuals above the higher threshold who come within the scope of the equal wrongness thesis). This account necessarily revises common sense morality in an effort to achieve greater consistency. It departs from common sense morality by assigning an intermediate and variable status to very young children, many of the radically impaired, and some higher nonhuman animals (It is, of course compatible with the view that, by virtue of their potential to become persons, very young children, and even fetuses below the lower threshold, might be subject to certain protections that might not apply to animals or some radically impaired human beings whose psychological capacities and potential are comparable to those of a young child. For what is done to a human individual as a fetus or as a young child can affect that individual adversely after she becomes a person and is above the higher threshold). The three-tiered account implies that the moral status of very young children increases as their psychological capacities develop and mature. As they become increasingly psychologically substantial, the degree to which they are inviolable increases correspondingly. Radically impaired human beings and some higher nonhuman animals have roughly the same degree of inviolability as a young child at the corresponding level of psychological development (apart from the point about potential noted parenthetically above). The difference is that a young child's status will continue to increase until it reaches the higher threshold, whereas that of a radically impaired human being or a higher nonhuman animal will remain at

the level corresponding to the level of psychological capacity at which that individual's development is permanently arrested.

The three-tiered account offers at least partial responses to some of the objections to the equal wrongness thesis noted above. In particular, it does not hold, as the two-tiered account does, that there is an enormous moral difference between those who are our moral equals and all those who are not. By contrast with the implication of the two-tiered account, the three-tiered account implies that when a child becomes our moral equal, he or she does not undergo a radical change of moral status but becomes only incrementally more inviolable. The change is only one of the degree of moral status rather than kind of moral status and is thus proportionate to the psychological change that underlies the change in status. The three-tiered account accepts, in other words, that different individuals, human and nonhuman, are varying distances from equality with persons who are above the higher threshold and that immature human beings approach equality gradually, by degrees, as they mature psychologically.

It also mitigates the problem that the exclusion of all animals from equality seems to require the exclusion of some human beings, including very young children, as well. The three-tiered account's solution will not satisfy common sense morality, since it too locates some human beings below the higher threshold. But at least it places some of those human beings above the lower threshold, assigning them an intermediate status, along with some animals with comparable psychological capacities.

The three-tiered account does, however, raise new questions. I have assumed that the two-tiered account rules out the sacrifice of a person above the threshold of respect even for the sake of an indefinite number of individuals below it. But matters are more complicated if there are two thresholds and the individuals between them are recognized as having variable degrees of worth and inviolability. It seems that those between the thresholds are in principle sacrificeable both for the sake of those above the higher threshold and for others between the thresholds. But could an individual above the higher threshold be sacrificed for the sake of some number of individuals between the thresholds? It seems that if there is to be a higher threshold at all, it ought to be sufficiently significant to protect those above it from being sacrificed for the sake of those below. Yet because the difference in psychological capacity between a person just barely above the higher threshold and an

individual just barely below it is very slight, it also seems implausible to suppose that a person just barely above this threshold cannot in principle be sacrificed for the sake of any number of individuals who are just below it. And yet it also ought not to matter that this person's psychological capacities are barely sufficient to get him above the threshold, since all those above the threshold are supposed to have the same maximum inviolability. According to the three-tiered account, this person has the same moral status as the person or persons who have the highest psychological capacities.

While the three-tiered account addresses some of the objections to the equal wrongness thesis, there are some that it does not address at all. And it may well be that some of these other challenges cannot be answered. If that is so, we must either accept the implications to which these challenges call our attention or else abandon the equal wrongness thesis, which is a significant element in common sense, liberal egalitarian morality. What would the implications of the latter option be?

One possibility would be to extend the principles that, according to the two- and three-tiered accounts, govern the treatment of individuals below the lower threshold so that they instead govern the treatment of all individuals. These principles are concerned with respect for interests. This option would involve repudiation of all constraints on action other than those imposed by considerations of consequences.

Another possibility would be to embrace a fully gradualist account of moral status, one without significant thresholds. The treatment of individuals with the lowest psychological capacities would not be governed by constraints at all. The treatment of those with only slightly higher capacities would be governed by very weak constraints. And beyond that the treatment of any individual would be governed by constraints the strength of which would be proportionate to the degree of development of the individual's relevant psychological capacities. The whole of morality, in other words, would have the same structure as the area between the lower and higher thresholds in the three-tiered account.

Such a view might hold that the killing of an individual is more seriously wrong the higher that individual's relevant psychological capacities are. Or it might hold instead that the degree to which it is wrong to kill an individual is determined by weighting that individual's interest in continuing to live by the level of his or her psychological capacities. That, it seems, is the way the view would most plausibly treat the infliction of nonlethal harms: an individual's

interests would in general be weighted for that individual's level of psychological capacity.

This sort of elitist view is profoundly counterintuitive. It would also be difficult to apply any scheme for weighting interests except by reference to a scale of psychological capacity with an upper as well as a lower bound; yet there is no determinable upper limit to human psychological capacity.

There are, however, certain common intuitions that tend to be suppressed by liberal egalitarian morality that might be captured by an alternative form of gradualist morality. It is a recurrent theme in the history of human thought, powerfully articulated even in the biblical books of *Job* and *Ecclesiastes*, that it is not only unfortunate but also unjust when the virtuous suffer while the vicious flourish. The sense that the distribution of good fortune ought to favor the virtuous is, I suspect, one of the factors that prompt people to believe that there is an afterlife. For people find it intolerable to suppose that the injustices of the life we know will remain forever unrectified.

Recall the earlier suggestion that many people accept that there would be a moral reason to save a 20-year-old rather than a 90-year-old if only one could be saved, or to kill a 90-year-old rather than a 20-year-old if circumstances make it permissible to kill one or the other. Suppose now that we alter the cases, making the first a choice between saving a virtuous person and saving a murderer of the same age (a murderer who has been punished according to his desert and is no longer dangerous but who feels no remorse), and the second a choice between diverting the train onto a track on which a virtuous person is trapped and diverting it onto a track on which the unrepentant murderer is trapped. When I ask my students about these cases, they are unanimous in thinking that one ought morally to save the virtuous person and, in the second choice, kill the murderer. Yet there is no sense in which the murderer is *liable* to be allowed to die, or to be killed. He has already been punished for his previous offense, so we may suppose that he no longer deserves to be harmed (if indeed he ever did), and we can also suppose that in both cases he is no more responsible than the virtuous person for the threat that people face. If there is a moral reason to favor the virtuous person in these cases, it therefore seems that it must derive from a comparative evaluation of the each person's overall moral worthiness, whether this is a function of the person's deeds or of his dispositions or character, or both.

The belief that differences in people's moral worth are relevant in these cases is, I suspect, both more common and stronger than the corresponding belief in the relevance of differences in the amount of good life that people would lose by being killed or allowed to die. To consult one's own intuitions, one might imagine further variants of the cases involving a choice between the life of a young but quite wicked person (who would happily cause serious harm to innocent people if he could succeed with impunity, but is sufficiently prudent to be predictably deterred by the threat of punishment) and an elderly but saintly person. If we assume that effects on third parties would be roughly equivalent in both cases, many of us would think it morally preferable to save, or to avoid killing, the elderly virtuous person.¹⁶

Next recall the seventh challenge to the equal wrongness thesis listed earlier. According to that challenge, if a certain consideration has moral significance in choices of whom to save, as well as in choices among killings when it is permissible to kill *someone*, then there is a presumption that that consideration should also be significant in comparisons between *wrongful* killings. Suppose that this challenge cannot be answered and that factors that are relevant in choices of whom to save and choices of whom to kill when it is permissible to kill someone are also relevant to determining the degree to which a wrongful killing is wrong. In that case, our intuitions seem to commit us to the view that the higher a person's overall moral worth (as determined by what she has done or by what her moral nature would lead her to do in relevant circumstances), the more seriously wrong it would be to kill her, if other factors remain constant.

Such an inegalitarian view seems less repellent than a view that would make the degree to which a killing is wrong depend on the level of development of the victim's psychological capacities generally. And, as I suggested, it also seems less implausible than the claim that the wrongness of killing varies with the strength of the victim's interest (or time-relative interest) in continuing to live. But it is nonetheless offensive to our liberal egalitarian commitments. It is, moreover, unclear what any view that makes the wrongness of killing a function of the psychological or moral nature of the victim would

¹⁶ It is possible to see such a view as an extension of plausible principles of moral liability. For arguments that indicate how we might be led in the direction of such a view by a series of small extensions of a plausible account of liability, see Jeff McMahan, "Self-Defense and Culpability," *Law and Philosophy* 24 (2005), pp. 760–765.

imply about the infliction of lesser or nonlethal harms. If the suggestion were that we should weight different people's interests in avoiding harms of all sorts for the degree of their psychological capacity or their moral virtue or worthiness, the implications would be seriously counterintuitive.

It is hardly necessary to express the further thought that readers will surely have already had – namely, that it would be dangerously invidious to give public expression to a view that accords a higher degree of moral inviolability to people with higher psychological capacities or a worthier moral nature. Even if such a view were true, it is virtually certain that if it were widely espoused and recognized as true, it would then be distorted or otherwise abused in efforts to justify the unjustifiable.

All this leaves me profoundly uncomfortable. It seems virtually unthinkable to abandon our egalitarian commitments, or even to accept that they might be justified only in some indirect way – for example, because it is for the best, all things considered, to treat all people as equals and to inculcate the belief that all are indeed one another's moral equals, even though in reality they are not. Yet the challenges to the equal wrongness thesis, which is a central element of liberal egalitarian morality, support Mulgan's skepticism about the compatibility our all-or-nothing egalitarian beliefs with the fact that the properties on which our moral status appears to supervene are all matters of degree. It is hard to avoid the sense that our egalitarian commitments rest on distressingly insecure foundations.¹⁷

Department of Philosophy
Rutgers University
New Brunswick, NJ, 08903, USA
E-mail: McMahan@Philosophy.Rutgers.edu

¹⁷ I am deeply grateful to Nir Eyal, Christopher Knapp, Rahul Kumar, Kasper Lippert-Rasmussen, and Carlos Soto for perceptive written comments on this article, and to Agnieszka Jaworska for unusually helpful discussion.