

## *Critical Notice*

JEFF McMAHAN. *The Ethics of Killing: Problems at the Margins of Life*. Oxford: Oxford University Press 2002. Pp. 540 + ix.

There is no moral belief that is more universal, stable, and unquestioned, both across different societies and throughout history, than the belief that killing people is normally wrong. Yet 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 commonsense beliefs about the morality of killing. (McMahan, 189)

In this exceptional new book, Jeff McMahan sets out to provide such an account. Along the way, he offers nuanced and illuminating accounts of personal identity, human nature, the badness of death, the wrongness of killing, the rights of animals, abortion, and euthanasia. This book is a major contribution to both moral theory and applied ethics, and makes a strong case for the relevance of the former to the latter. It is also beautifully written and a joy to read.

The practical relevance of a philosophical analysis of killing lies in the fact that 'an understanding of why killing is normally wrong should help us to identify the conditions in which killing may not be wrong' (189). In particular, McMahan seeks an account of killing that makes sense of contested terrain at the margins of human life (abortion and euthanasia) and also the morality of the taking of non-human lives.

McMahan's methodology is a relatively standard one in contemporary moral philosophy. He combines conceptual analysis with judicious appeal to widely shared intuitions, and tests the results against a variety of counterexamples, both real and imagined. His use of this method is exemplary, as is his defence of it: 'I am going to assume, in this book, that, unless they can be explained away as obvious products of collective self-interest, exploded metaphysics, factual errors, or some other dis-

crediting source, our common moral intuitions should be treated as presumptively reliable, or as having some presumptive authority' (238).

This is a long, dense book, overflowing with examples, arguments, and counterarguments. Insofar as there is a unifying theme, it is McMahan's defence of a two-tiered account of the wrongness of killing, the view 'that morality, and in particular that area of morality concerned with killing, is divided between a region concerned with interests and well-being and a region concerned with respect' (260).

Once this division is in place, McMahan's 'main purposes are to defend the permissibility of certain killings of beings below the threshold, and to show that some killings of persons are not wrong at all because they are compatible with what is required by respect' (251). In other words, McMahan mounts a partial defence of abortion and euthanasia.

The remainder of this review aims to summarise the main points of each of McMahan's chapters, to show their interconnection, and also to raise a few doubts and concerns, not so much about the details of McMahan's account, but about his enterprise as a whole. In a brief review, it is not possible to do justice to all the topics covered in the book. In particular, I skip lightly over McMahan's initial discussion of what we are, as well as much of his more applied discussion, to concentrate on the central notion of killing.

## I Identity

To kill *x* is to cause the death of *x*. Intuitively, killing is wrong because death is bad. To know what counts as death for 'one of us' (McMahan's neutral term), we must know what we are, and what possible future events would count as killing one of us rather than changing him or her. To take a simple contrast, if you chop off my head, you kill me, whereas if you chop off my finger, you do not. Yet both my head and my finger are parts of me. Presumably my head does something that is more essential to my identity. But what exactly?

McMahan defends an '*embodied mind account* of personal identity, according to which the criterion of personal identity is physical and minimal functional continuity of the brain' (69). Me-today is the same person as me-yesterday because the same brain supports a consciousness today that is connected to the consciousness it supported yesterday.

McMahan argues that there is a close connection between egoistic concern and personal identity. If I am egoistically concerned for some future person at all, he will typically be me. (Hypothetical cases of division provide an exception here, as McMahan follows Derek Parfit in holding that I can be egoistically concerned for each of two future people,

even though neither is identical to myself.) However, identity is not all that matters, even for egoistic concern. For instance, psychological connectedness, beyond the minimum needed for continuity of consciousness, is not relevant to the determination of personal identity. However, my degree of psychological connection to a future person is a key factor in determining whether or not I feel egoistic concern for her.

To illustrate this connection, McMahan asks us to 'imagine the prospect of becoming like a god ... of becoming vastly more intelligent and developing a vastly richer and deeper range of emotions, including emotions of which one cannot now form any conception' (321). He argues that 'even if the transformation would be identity preserving and would lead to a state that would be clearly superior to one's present state, it would be too much like becoming someone else to be very desirable from an egoistic point of view' (322). Even if the future god would be me, she is so different from me that my concern for myself leads me to prefer a less desirable future state more continuous with my present one.

A crucial issue for the embodied mind account is 'the relation between the mind and the human organism' (89). Is my existence temporally coextensive with the existence of my organism? Or could that organism either exist before I do, or continue to exist after I am gone? The embodied mind account answers both questions in the affirmative, as an organism can live before (and after) it supports a conscious mind. These answers are key elements of McMahan's defence of abortion and euthanasia, as we shall see.

## II Death

Following much contemporary philosophy, McMahan assumes that when we die we cease to exist. This is not mere stipulation. It follows from his account of what we are. 'Unless there is reason to suppose that our brains will be resurrected, the Embodied Mind Account of Identity seems to rule out the hope of an afterlife' (98). McMahan thus sides with the traditional Christian view that the physical resurrection of this very body is necessary for any personal after-life. Proponents of alternative views of the afterlife will argue that, if the embodied mind account is inconsistent with life after death, then so much the worse for that account. They might argue that the embodied mind account only fits our usual practices, and thus only appeals as an intuitive account of 'what we are' if one has already decided that there is no afterlife.

Death is thus bad because it is the end of our existence. McMahan argues that, in practice, our interest in the badness of death is always comparative. The main question is always 'how bad it is for a person to die this death now rather than a later death from a different cause?' (104).

McMahan contrasts two accounts of the badness of death. On the 'life comparative account', we compare the life that is ended by death with the life the person would have lived if they had not died then. The badness of that death is a function of the amount of good the person loses by dying. The 'time-relative interest account' links the badness of death for a person at a particular time to the strength of their interest at that time in continuing to live.

The two accounts generally go together, but they can come apart. Suppose a person's future self would only be loosely psychologically connected to her present self. Her present interest in continuing to live is thus comparatively weak. Recall the example of being transformed into a god. Suppose I die the moment before I would have been miraculously transformed. Because the future god-like person would only be very loosely connected to me, my time-relative interests may not be significantly thwarted by my death. According to the life comparative account, however, my death is about as bad as any death could be, as the total value of the life I lose is enormous.

If I have no strong interest in becoming a god, then it is no *misfortune* for me not to be transformed into a god. If the only alternative to death is something whose loss is no misfortune to me, then death is not (comparatively speaking) much of a misfortune either. (Of course, it *is* a misfortune that these are my only options, but that is another story.) McMahan concludes that the time-relative account is superior.

The two accounts might be used for different purposes. For instance, we might ask whether it is worse for person *x* to die at *t*<sub>1</sub> (as a newborn infant) or at *t*<sub>2</sub> (as an adult). Obviously enough, it is better for *x* if he lives until *t*<sub>2</sub> and then dies, as this gives him a much better life. The life comparative account is appropriate for this purpose. Alternatively, suppose *y* (a newborn infant) and *z* (a 35-year-old in perfect health) both die at *t*. Who suffers the greater misfortune? In this case, the time-relative account seems more appropriate. Although *z* has lived a better life than *y*, *z*'s death is a much greater misfortune for *z* than *y*'s death is for *y*. A newborn infant has her whole life ahead of her, and thus in one sense loses most by dying. Yet that future life is only very weakly psychologically connected to her present state. Her present interest in living that future life is thus much less than *z*'s interest in continuing his life.

The gap between these two accounts of the badness of death is the second key to McMahan's defence of abortion.

### III Killing

McMahan's account of killing, the centrepiece of the book, begins with a deceptively simple claim. 'It is uncontroversial that the killing of an animal is normally less seriously wrong than the killing of a person' (190). The first challenge for any account of killing is to accommodate this simple claim, as well as a variety of more complex claims about the comparative badness of killing different human beings.

One obvious explanation is that the wrongness of killing depends on the species of creature killed. Killing a human being is worse than killing, say, a chimpanzee, because the two belong to different species. One of McMahan's principal aims is to undermine this 'species membership account.' He invites us to 'imagine a spectrum of possibilities, animal to human, with the change of a single gene at each stage. It would be absurd to suppose that the moral status of any individual in the spectrum would be determined by how many, or what proportion, of its genes were human or were taken from a human being' (213).

McMahan uses the fact that such a spectrum can be constructed to show that species membership itself cannot be the basis of the great moral difference we see between animals and (normal) human beings. However, such spectra can be used to make a more general point. When any physical property is presented as underlying the moral difference between animals and humans, we can construct a spectrum where *that* property changes minutely at each stage. Once again, each individual in the spectrum is morally indistinguishable from those before and after. This seems, on the face of it, to undermine the claim that the first and last points differ in a morally significant way. (I return to this more radical possibility in Sections VI and VII below.)

McMahan's own account bases the wrongness of killing *x* directly on the badness of *x*'s death. This account faces an immediate challenge, as 'whatever we take to be the range of psychological capacities that differentiate us morally from animals, there are some human beings whose psychological capacities are no more advanced than those of certain animals' (204). This suggests that severely intellectually disabled humans should be treated exactly the same as animals with the same psychological capacities. Indeed, McMahan endorses this conclusion, arguing that 'we should treat animals no worse than we believe severely retarded human beings with comparable capacities should be treated by moral agents who are not specially related to them' (227). If our intuitions regarding animals and similarly capable humans diverge, McMahan argues that we should reach convergence by upgrading the moral status of animals far more than we downgrade the moral status of severely retarded human beings.

Killing is wrong because death is bad for us. The obvious next step is to equate the wrongness of each particular killing with the badness of the death that occurs. If killing is wrong because death is bad, then presumably any particular killing is as wrong as its associated death is bad. However, this simple equation is very problematic. According to both accounts of the badness of death, the badness of different deaths varies widely. In particular, death is often less bad for a very elderly person than for someone in the prime of life, as the relevant alternative to this particular death is another death in the not-too-distant future. This would imply that killing the former is less wrong than killing the latter. As McMahan concedes, almost none of us believe this.

In response, McMahan offers a two-tiered account of the wrongness of killing. This is embedded in a two-tiered account of morality in general, based on 'a threshold on a scale that measures the psychological capacities that relevantly distinguish persons from animals' (245). In developing his two-tiered account, McMahan draws explicitly on Warren Quinn's division of morality into a morality of humanity and a morality of respect. (McMahan renames the former the 'morality of interests,' to emphasise the fact that our obligations regarding animals fall in this realm [245].) *Persons* are above the threshold, while creatures who are not persons are below it. This distinction, in turn, is to be cashed out in terms of certain higher psychological capacities that distinguish persons from most or all forms of animal life, and from some forms of human life as well.

Killing an animal is wrong primarily because of the badness of death for the animal, and the wrongness will be proportional to the badness. Killing an adult human being, by contrast, is wrong because it fails to respect that person. As the worth of a person is unrelated to the badness of her particular death, so is the wrongness of killing her. As each person has equal worth, each killing is equally wrong. (This approach also opens the possibility of killings that are not wrong at all, because they *do* respect the person's worth. This is the basis of McMahan's defence of (some forms of) euthanasia.)

At this point McMahan raises the spectre of a spectrum argument undermining his own theory. 'Morally, the gap between those above the threshold and those below it is immense ... yet it seems that each of us must cross that line with a single step' (261). Obviously enough, we can imagine a spectrum running from the lowest animal to the smartest human. We cannot say where in this spectrum 'personhood' emerges. Doesn't this show that the moral gap between persons and animals cannot be as significant as McMahan's theory requires?

McMahan has two responses. The first is to note that 'we are also unable to specify the precise moment that a person becomes three feet tall; but that alone does not necessarily convince us that there is no such

moment' (261). Unfortunately, this is not altogether satisfactory. Derek Parfit's original use of the spectrum argument to defend a Reductionist account of personal identity is instructive here. (See Derek Parfit, *Reasons and Persons* [Oxford: Oxford University Press 1984], Part Three.) Parfit's spectrum does not show that there are no persons, or that we cannot make sense of claims such as 'x is the same person as y.' Rather, it aims to show that the concept 'person' cannot bear the weight assigned it by certain Kantian approaches to morality. Similarly, suppose someone suggested that 'being three feet tall' was a moral distinction as significant as the distinction between persons and animals that McMahan requires. The fact that height is continuous does seem to effectively undermine *that* claim.

McMahan's second response is that, in practical terms, our uncertainty over the threshold need not concern us, as almost all practical cases of abortion or euthanasia occur a long way from the threshold, wherever that threshold might be located. I consider this defence in the final two sections.

We turn now to McMahan's detailed discussion of abortion and euthanasia. The basis of both is the possibility of divergence between my organism, my continued existence as myself, and my status as a person. Abortion and euthanasia will be permitted if what is killed is an organism that does not embody a mind, and they may be permitted if what is killed is one of us who is not a person.

#### IV Beginnings

Under the embodied mind account, the first morally significant point in a human organism's life is the point at which the foetal brain develops the capacity to generate consciousness. Only then does the organism become one of us. McMahan notes that 'Most neurologists accept that the earliest point at which consciousness is possible is around the 20th week.' Therefore, if we define an 'early abortion' as one occurring prior to 20 weeks, it follows that 'an early abortion does not kill anyone; it merely prevents someone from coming into existence' (267). This is highly significant in practice, as 'approximately 99 percent of all abortions in the United States are performed prior to 20 weeks' (268).

McMahan's theory thus easily permits early abortion. By contrast, 'a late abortion [one performed after consciousness develops] involves the killing of someone rather than merely something. It involves the killing of one of us' (269). On the other hand, if we favour the time-relative account of the badness of death over the life comparative account, late abortion does not involve a death that constitutes a great misfortune for the creature who dies. More significantly, although the developed foetus

is 'one of us' in the sense that it possesses the capacity for consciousness, it 'must fall outside the scope of the morality of respect, for it clearly does not possess the capacities that distinguish persons from animals' (263). A developed foetus is 'one of us,' but not in the sense that really matters. Late abortion is no more wrong than killing a (similarly psychologically sophisticated) animal.

If developed foetuses are not persons, then this affects not only the degree of wrongness of killing them, but whether killing is wrong at all. For instance, we deal with epidemic diseases among animals by slaughtering some to save others, while we would never contemplate such a solution for adult humans. If developed foetuses are on a moral par with animals, then it is also acceptable to sacrifice them to save the lives of others.

Not everyone will welcome this conclusion. Unfortunately for McMahan, there is much worse to come. Given his account of the morally significant points in a human being's life, everything he says about developed foetuses applies to newborn infants. This raises a general problem for McMahan, as he acknowledges that 'most people draw a sharp moral distinction between abortion and infanticide' (339). Indeed, 'it is virtually impossible to believe that infanticide is morally indistinguishable from abortion' (342).

Many moral theorists struggle with the contrast between infanticide and abortion. However, McMahan's view has particularly 'shocking' implications here. Here is 'the worst-case' example: 'According to the view I have developed, it ought to be permissible, if other things are equal, to sacrifice [a] newborn orphaned infant in order to save [three other] children' (355).

This implication follows directly from the fact that, like developed foetuses or animals (but unlike persons), the killing of newborn infants is only wrong because it violates their interests. If killing best promotes interests overall, there is no objection to it. McMahan admits that 'most people find this implication intolerable, and I confess that I cannot embrace it without significant misgivings and considerable unease' (360). To defend the comparative intuitive appeal of his own approach, McMahan seeks to show that the leading alternatives are equally problematic.

If an abortion is performed, then the person the foetus would have become never exists. So, McMahan argues, an argument against abortion cannot appeal to any interest that person would have had, as never-existing interests count for nothing. We must focus on the actual interests of the foetus.

The most common alternative to McMahan's approach argues that the foetus has a special moral significance due, not to its actual abilities, but to its potential. In McMahan's own terminology, this amounts to the



claim that a foetus has a morally relevant interest in fulfilling its potential to become a person.

McMahan's response to this alternative begins with a distinction between two kinds of potentiality: identity-preserving and non-identity-preserving. To illustrate the contrast, imagine two ways someone might bring it about that my body becomes the location of a being with godlike intelligence. The first is to offer me a regime of training and drug therapy that would transform me into such a being. The second is to obliterate my personality altogether, replace the neural networks in my brain, and rearrange the new network to create a new super-intelligent being. If the first option is available, I have the identity-preserving potential to achieve godlike intelligence. If only the second option is available, then the future intelligence will not be me, and so my potential to be godlike is non-identity-preserving.

As the early foetus is not 'one of us,' it 'does not have the identity preserving potential to become a person. It has only the non-identity potential to become a person' (305). However, this kind of potential is nothing special. After all, 'any healthy pair of sperm and egg has the same non-identity potential that the early foetus has to become a person' (306). As a sperm and egg pair has no special moral right to become a human being, neither does an early foetus.

By contrast, a developed foetus *does* have the identity-preserving potential to become a person. I can say 'I was once a developed foetus,' while no-one can truthfully say 'I was an early foetus.' However, as even the developed foetus lies below the threshold for respect, 'we should be guided by the developed foetus's time-relative interests, and its time-relative interest in realising its potential to become a person is weak for the same reason that its time-relative interest in continuing to live is weak' (308). McMahan's view here follows from his preference for the time-relative account of the badness of death, together with a claim about the strength of a developed foetus's present time-relative interests. Because the foetus would be only weakly psychologically connected to the person it would become, and because (unlike, say, a person who has devoted his life to becoming a god) it has invested comparatively little in becoming that person, it has only a weak interest in fulfilling its potential to become a person. McMahan argues at some length that these claims hold true even if becoming a person is what would 'naturally' happen to the foetus.

McMahan concludes that neither kind of potentiality could grant the foetus a higher moral status than it enjoys under his theory. As a newborn infant has the same potential as a developed foetus, potentiality also cannot be used to explain the moral status of infants.

## V Endings

At the end of life, the significant feature of the embodied mind account is that, 'if we are not identical with our organisms, one of us can cease to exist even if his or her organism remains in existence and, indeed, even if it remains alive' (424). An organism can continue to live even after it has lost the capacity to support consciousness, while 'we must cease to exist when we lose the capacity for consciousness in a way that is in principle irreversible' (423). This leads McMahan to conclude that, for moral purposes, 'the notion of brain death does not do justice to our intuitions about death' (427), as the brain can continue to live long after consciousness is irrevocably destroyed. Instead, 'the appropriate criterion must be the destruction of those areas of brain in which consciousness is realised. The problem is that as yet no-one knows exactly which areas those are' (439).

As with abortion, this leads to a bifurcation of cases of euthanasia. If a human organism has irretrievably lost the capacity for consciousness, then it is no longer one of us. It thus has the same moral status as a human corpse; or, in some respects, the same moral status as an early foetus. If a human organism still possesses the capacity for consciousness, there are two separate cases to consider. The first is when a person wishes to die, while the second is when one of us is no longer a person.

In the first case, McMahan defends the rationality of suicide, and argues that, in such a case, euthanasia is consistent with the requirement of respect for the person killed. His target here is the Kantian view that killing a person is always a violation of respect. McMahan offers a forceful response to the Kantian view: 'If told by a Kantian that he must not sacrifice his rational nature for the sake of his good, a person who desperately wants to die could reasonably conclude that he was being subordinated to the value of his rational nature' (479).

The second case corresponds to the developed foetus or newborn infant: one of us who is not yet a person. However, the conscious no-longer-a-person differs in one crucial respect from the conscious not-yet-a-person, as we can appeal to the preferences and values the person herself previously had. This creates a potential for conflict, as when a person asks not to be resuscitated if dementia takes hold, but then, in the grip of dementia, expresses a wish to continue living. In McMahan's terminology, her (past) autonomous preference as a person must be balanced against her time-relative interests as a no-longer-person. He admits that, in general, 'it is not obvious what respect for a person requires when it is in her interest to live, but her autonomous preferences is to die' (498). However, McMahan's conclusion is that, in most cases, 'the demented patient's physicians ought to give priority to the earlier part of her life, or to her earlier self. They ought, in other

words, to allow her to die. Because the earlier part of the life is overwhelmingly the dominant part, it should have priority' (502). Part of the explanation for this is that, 'there is often reason to believe that many of an individual's values persist well into dementia, even when the individual loses the capacity to be guided by those values in her behaviour.... It may be that the values remain and what she has lost is the ability to understand that her continued existence is incompatible with them' (503).

In both abortion and euthanasia, the interests of a person trump those of one of us who is not a person, let alone the interests of a being who is not one of us at all. The priority of persons is thus the key to the resolution of conflicts between a woman seeking an abortion and her developed foetus, and between a demented patient's past and present preferences.

## **VI Those of us who are not yet persons**

The basis of McMahan's moral theory is a distinction between five types of human being: (a) a human organism who is not yet one of us; (b) one of us who is not yet a person; (c) one of us who is a person; (d) one of us who is no longer a person; and (e) a human organism who is no longer one of us. The most significant difference is between persons and human organisms who are not persons. Accordingly, our attention is focused on the two boundaries of this concept: the gray areas where personhood emerges and disappears. I will concentrate on the former.

A human organism becomes one of us when it develops a brain capable of supporting consciousness. This happens very early in its development. Even if we do not know exactly when this threshold is passed, we at least know precisely what we are looking for. The transition to personhood is much more problematic. It occurs much later, and over a longer period. Nor can we even agree on what we are looking for, as different accounts of the moral basis of personhood have different implications for when one of us becomes a person. Few moral decisions concern human organisms in the process of becoming one of us, whereas many very difficult moral decisions relate to those of us who are in the process of becoming persons.

We begin with a common pair of intuitions. Many people feel that, while killing an early foetus is not wrong, it is very wrong to injure a foetus in a way that harms the resulting person. For instance, suppose I stand to inherit a considerable fortune unless I one day have a nephew who has children. I decide to protect my inheritance by rendering my future nephew infertile. I know it would be wrong to wait till he is fifteen years old, and then sterilise him, as this would clearly be a failure of respect for a person. So I decide to sterilise the early foetus that will one

day grow into my nephew. I defend myself as follows. If killing that foetus would not be wrong, how can it be wrong to harm it to a lesser degree?

As McMahan notes, this conclusion is unpalatable. 'If the time-relative interest account implies that it would be less objectionable to cause a person to be sterile by injuring him when he was a foetus than to cause him to be sterile by injuring him at age 15, this is seriously damaging to the account's credibility' (282). After all, the only difference is that the harm occurs earlier. How could this make things *better*?

McMahan's solution is to appeal, not to the early foetus's present time-relative interests, but to the (future) time-relative interests of the resulting person. 'The important consideration is whether one's action frustrates a time-relative interest; it does not matter whether the act is done before the time-relative interest exists' (283). My present action is wrong because it harms a future person. To adapt an analogy offered by Derek Parfit, my action in sterilising the early foetus is morally analogous to planting a landmine in a forest that no-one will enter for at least one hundred years. This is wrong because it will (predictably) harm a person, even though the person who will be harmed does not exist at all at the time of the action.

This way of thinking certainly solves this particular puzzle. Unfortunately, it gives rise to another counterintuitive result, as McMahan himself notes. Suppose my inheritance depends, not on my nephew not having children, but on his never becoming a person at all. Suppose our eccentric Kantian ancestor has established a rigorous tests of rational autonomy that must be passed by any inheritor. Unable to prevent my sister from conceiving, I harm her early foetus so that it never develops into a person.

Intuitively, my action in this case is at least as bad as before. Certainly, if I waited till my nephew was fifteen and *then* destroyed his capacity for personhood, this would be worse than rendering him unable to reproduce. Yet any theory such as McMahan's has great difficulty accommodating this intuition. Instead of thwarting existing (present or future) interests, I prevent those interests from arising at all. I have not harmed any future person, because the whole point of my action is to ensure that there is no future person. Accordingly, no-one is harmed by my action. My action is analogous to a decision to create a severely retarded human being rather than creating a person. Such a decision cannot be faulted for harming the future person who is never created. 'The act that we wish to condemn has the effect of preventing itself from coming within the scope of the morality of respect' (324).

McMahan explores various possible solutions here, all of which he finds unsatisfactory. He reluctantly concludes that 'I must end this section inconclusively.... There do not seem to be any considerations that

plausibly explain and justify our intuitions about these various cases' (327-8).

These problems stem from McMahan's insistence on dealing, below the threshold of personhood, only with actual interests, not with an individual's potential personhood or with species-relative norms of development. To explore this feature of his view, we now consider a range of cases between foetus and adult. There is no fixed point when one of us becomes a person. For an indefinite period, we are beings of an intermediate status in the process of becoming persons. McMahan largely put such beings to one side. (Although he does briefly address the issue of how the killing of such a being relates to a killing of a person, and suggests the possibility of an intermediate moral realm between those of interests and respect [265].) I shall now attempt to extend his account to cover these intermediate cases, and thus raise some additional difficulties.

A developed foetus has only a very weak time-relative interest in becoming a person. As the human being develops, this interest strengthens for two reasons. The psychological connections between successive time-slices of the human being increase in number, complexity, and moral significance. Her interest (today) in being alive tomorrow increases each day. Furthermore, the length of time remaining until the human being becomes a person *reduces* each day. So her connection to the person she will be grows stronger, as does her (present) interest in becoming that person.

Even if we confine ourselves to the morality of interests, killing an older child who is not yet a person is thus much more morally problematic than killing a developed foetus or a newborn infant. However, on its own, even the strongest time-relative interest seems insufficient to generate an intuitively appealing morality regarding children. A child who is not a person is not entitled to the respect due to persons. This raises several familiar problems, which are unfortunately worse in the case of older children than in the case of foetuses or newborn infants. These problems all relate to the implicitly utilitarian character of the morality of interests as presented by McMahan.

The first problem is replaceability. If all that matters about *x* is the satisfaction of *x*'s interests, then there is no moral objection to replacing *x* with another creature whose interests can be satisfied to a higher degree. A happy sheep can be costlessly replaced, from a moral point of view, with a happier sheep. A notorious feature of Peter Singer's discussion of newborn human infants, similar to McMahan's view, is that he treats them as replaceable. An older child who is not yet a person, and thus not entitled to respect, would also be replaceable. It might be much harder to find a comparable replacement, given the complexity of the child's interests, but there is no intrinsic moral objection to doing so.

Intuitively, this conclusion seems even less attractive with older children than with infants, foetuses, or animals.

A second worry here is the threat of an analogue of Derek Parfit's infamous 'Repugnant Conclusion'. Within the morality of interests, the interests of different creatures are weighed against one another. Even if the time-relative interests of a developed foetus or an animal count for comparatively little, a sufficiently large number of such interests should have more value than the time-relative interests of any human being who is not yet a person. So we can construct imaginary cases where we would be required to sacrifice the life of an older child in order to save the lives of a sufficiently large number of sheep. The implications for public morality, especially concerning the allocation of scarce resources between competing needs, threaten to be quite radical.

Such repugnant conclusions also threaten to overwhelm the value of persons. The most plausible solution is to argue that the value of a person's life has lexical priority over the lives of animals who are not persons. No amount of the value contained in lives of the latter sort would outweigh the loss of a person's life. Such lexical differences make sense when we are dealing with creatures of radically different kinds, such as a rational autonomous adult human being versus a sheep. However, it is much harder to defend lexical differences in value *within* the realm of the morality of interests. At what point does the child's life suddenly become incomparably more valuable than anything that went before? (The fact that childhood shades imperceptibly into personhood also undermines the idea of a lexical gap between persons and animals, as we shall see in the next section.)

A third problem is that, by McMahan's own account, the distinction between doing and allowing is particularly difficult to justify in relation to the morality of interests. This suggests that the kind of sacrifice that is permitted for newborn infants will also be permitted for older children who are not yet persons. We could find ourselves permitted (perhaps even obliged) to sacrifice one ten-year old to save several others.

One solution to all these worries is to argue that older children are entitled to a certain kind of respect, which differentiates them from both the morality of interests and the morality of respect. It is inappropriate to treat an older child as if they were an adult, especially when such treatment will predictably prevent the child ever becoming a person. (For instance, we should not respect a ten-year old child's desire to become addicted to marijuana.) On the other hand, a creature who has a sufficiently strong time-relative interest in becoming a person is entitled to a certain kind of quasi-respect.

The challenge is to explain how quasi-respect operates, how it differs from the respect due to persons, and what exactly it is about older children that entitles them to quasi-respect. In particular, McMahan

argues strongly against treating older children differently from animals on the basis of their potential to become persons. Suppose we develop the technology to confer rational autonomy on a sheep. Intuitively, we believe our obligation to enable a child to develop rational autonomy is incomparably stronger than any obligation we might have to use this technology.

One solution, drawn from remarks McMahan himself makes in a related context, would be to argue that *parents* have particular kinds of obligations to their own children (376). In the actual world, these obligations are owed to human children, not because they themselves are human beings, but because they are the children of persons. If a sheep were the child of a person, then that person would have the same obligations to that sheep as a human parent has to her human child.

Aside from the general difficulty of providing a rational ground for, and explication of, parental obligations, this solution faces the following problem. It implies that children who do not have parents, or children whose parents are not (for whatever reason) themselves persons, are not entitled to demand (from other persons) the care needed to enable them to develop their capacity for personhood. Unrelated persons would have no obligation to provide such care rather than funding the development of the 'sheep-to-person-converter.' Once again, this conclusion does not seem entirely satisfactory.

My reason for focusing on the case of older children who are not yet persons is to draw attention to some general problems with McMahan's theory. I explore the implications of these in the next section.

## VII Problems for Naturalising Ethics

I suggested earlier that the threat of the spectrum argument has more severe implications for McMahan's view than he acknowledges. His view is not alone. Similar spectrum arguments threaten many other theories in this area. If *all* possible theories faced a spectrum argument, then perhaps it would be no objection to McMahan. Unfortunately, there are theories that do not. More unfortunately, for those of us committed to exploring the implications of naturalistic (or, at least, not overtly supernaturalistic) ethics, these include the very theories McMahan puts to one side for their 'theological' basis.

In recent moral philosophy, the spectrum argument owes its prominence to Derek Parfit's discussion of Reductionist accounts of personal identity. Parfit uses a series of spectrum arguments to show that, if personal identity supervenes on psychological and physical continuity, then the identity of persons cannot have a moral significance of an altogether greater kind than the moral significance of those underlying

properties. To oversimplify, Parfit uses a Humean account of the basis of personal identity to undermine a Kantian approach to morality.

The problem this poses for McMahan's project is as follows. 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, as demonstrated by the ever-present threat of spectrum arguments.

McMahan suggests that, because most (if not all) controversial moral issues occur at the uncontroversial ends of the spectrum, the spectrum argument is a theoretical puzzle we can typically put to one side in practice. It would certainly make life much easier if this were so. However, several things count against this comforting interpretation. The first is that, in practice, the implications of McMahan's theory regarding infanticide threaten to be extremely radical. In this case, McMahan seems to give an unacceptable answer to an *uncontroversial* moral question. The second problem is that, as we saw in the previous section, McMahan's approach faces difficulty accounting for a range of moral issues concerning our treatment of children who are not yet persons. These issues arise squarely on the boundary of personhood. Indeed, they arise because of our inability to precisely locate that boundary. A related problem is that, once we focus on the whole spectrum of cases between conception and adulthood, it becomes harder to justify the claim that the value of persons differs in kind from the value of animals or fetuses. Where exactly in childhood does this new kind of value emerge? Finally, it is by no means obvious that all moral theories do share all McMahan's difficulties here. Accordingly, we must at least consider an alternative diagnosis.

One might see the spectrum argument, not as a theoretical puzzle that any moral theory must face sooner or later, but rather as an artefact of McMahan's desire to retain two separate, and ultimately incompatible, worldviews: the Humean and the Kantian. On this reading, the spectrum argument indicates the presence of a 'discrediting source' of precisely the kind that, as McMahan himself admits, should cause us to question the reliability of our intuitions, and thus to question the strategy of seeking to reconcile them.

For the sake of argument, suppose we accept, contra McMahan, that the spectrum argument does show a deep tension in our intuitions. The most obvious way to resolve that tension is to abandon one or other of the two conflicting worldviews. The approaches of Hume, Parfit, and Kant are instructive here. Hume himself clearly thought that the consistent adoption of his naturalistic approach would have radical consequences for ethics. Parfit argues in a similar vein, when he suggests that the acceptance of Reductionism tips the scales, perhaps decisively, in



favour of Consequentialism over any Kantian approach to ethics. Kant himself equally clearly agreed. Only a radical departure from the naturalistic standpoint, at least for the purposes of practical reason, could enable us to salvage liberal concepts of personhood, autonomy, or respect. Kant acknowledged that the only way to save our intuitions is to bring in a non-natural set of 'ideas of reason', including not just autonomy, but also God and immortality.

Consider, for instance, the key claim that all persons have equal worth, and thus that the killing of any person is equally wrong. The notion of equal worth sits uneasily with any naturalistic picture of what gives rise to worth, as the underlying properties necessarily come in degrees. This is a perennial problem for naturalistic liberals, the dominant party in contemporary moral and political philosophy. Jeremy Waldron has recently suggested that, in the early classic liberal presentation of John Locke, the idea of equal worth is defended on explicitly theological grounds (Jeremy Waldron, *God, Locke, and Equality* [Cambridge: Cambridge University Press 2002]). Human beings have worth because we are created in the image of God. We all have *equal* worth because we are all equally created in the image of God. On this interpretation, modern liberals who take Locke's egalitarianism without his theology are like modern Kantians who want Kant's autonomy, liberalism, and equal respect without his (embarrassing) commitment to an afterlife presided over by a Christian God. The result is a foundationless set of intuitions. Unsurprisingly, the search for a coherent naturalistic foundation inevitably fails. If even our most considered intuitions are so unstable, then it is no longer clear that moral philosophy can (or should) be built upon them.

Jeff McMahan's book is a tour de force of contemporary naturalistic ethics. Defenders of the naturalistic project will cite it as the best evidence yet that the project is on the right track, yielding insights into a wide range of pressing topical issues. Critics of the project will argue that McMahan's inability to completely dissolve the spectrum argument is the best evidence yet that the naturalistic project must fail. No-one on either side can afford to ignore this book.

TIM MULGAN  
University of Auckland  
Auckland, New Zealand

