

## **On the very idea of criteria for personhood**

### *Abstract*

I examine the familiar criterial view of personhood, according to which the possession of personal properties such as self-consciousness, emotionality, sentience, and so forth is necessary and sufficient for the status of a person. I argue that this view confuses criteria for personhood with parts of an ideal of personhood. In normal cases, we have already identified a creature as a person before we start looking for it to manifest the personal properties, indeed this pre-identification is part of what makes it possible for us to see and interpret the creature as a person in the first place. And that pre-identification typically runs on biological lines. Except in some interesting special or science-fiction cases, some of which I discuss, it is human animals that we identify as persons.

## On the very idea of criteria for personhood

Meine Einstellung zu ihm ist eine Einstellung zur Seele. Ich habe nicht die *Meinung*, dass er eine Seele hat.

My attitude towards him is an attitude towards the<sup>i</sup> soul. I am not of the *opinion* that he has a soul. (Ludwig Wittgenstein, *Philosophical Investigations*, II, p.152, tr. G.E.M.Anscombe)

Love is conditioned by its object, but love also yields its object. (Gaita 1991: 122)

### 1. Persons: humanism, speciesism, criterialism

Any ethical outlook much like ours will take as central some *primary moral constituency* (PMC): some class of creatures who all alike, and all equally, share in the highest level of moral rights and privileges. Most philosophical ethicists use “person” to mean at least “member of the primary moral constituency” (whatever else they may also mean by “person”). In this paper, that is how I shall use the word too.<sup>ii</sup>

So what makes any creature a person, a member of the primary moral constituency? One view is what I shall call humanism:

**Humanism:** Being human is sufficient for membership of the PMC.

Humanism faces two very common objections. The first is that humanism is arbitrary discrimination, like racism, sexism, ageism; humanism, in Richard Ryder’s famous phrase, is *speciesism*. The second objection, which is really just the same idea taken a little deeper, says that humanism bases its demarcation of persons on a biological property which is morally insignificant. Membership of this or that species—just like membership of this or that race, gender, or age-group—is, we are told, a morally insignificant detail. It is not the kind of thing that we should expect any important moral distinction to be based on.

In theory these objections could be put by anyone who rejected humanism. In practice they are usually put by proponents of another theory of personhood, which I shall call criterialism:

**Criterialism:** Actual possession of the criterial properties is necessary and sufficient for membership of the PMC.<sup>iii</sup>

Here are four very well-known enunciations of criterialism<sup>iv</sup>:

I propose to use “person”, in the sense of a rational and self-conscious being, to capture those elements of the popular sense of “human being” that are not covered by “member of the species *Homo sapiens*”. (Singer 1993: 87)

Persons are beings capable of valuing their own lives. (Harris 1985: 16-17)

An organism possesses a serious right to life only if it possesses the concept of a self as a continuing subject of experiences and other mental states, and believes that it is itself such an entity. (Tooley 1972: 82)

[The six key markers of personhood are] (1) sentience... (2) emotionality... (3) reason... (4) the capacity to communicate... (5) self-awareness... (6) moral agency. (Warren 1997: 83-84)<sup>v</sup>

As here represented, the criterial view is an all-purpose<sup>vi</sup> view about what personhood is, and why it matters, which is undeniably simple (as I shall argue, much too simple). Despite this simplicity, it is obvious, when criterialism is stated as baldly as it is in these four statements, that it has radical implications. Most noticeably, it implies a sharp contraction of the circle of intuitive moral concern. If rationality is a necessary condition of counting as a person, a member of the primary moral constituency, then all children under a certain age will fail to count as persons. (This bullet is explicitly bitten by Tristan Engelhardt: “If being a person is to be [*sic*] a responsible agent, a bearer of rights and duties, children are not persons in a strict sense” (Engelhardt 1989: 120).) If self-consciousness is a necessary condition, then many mentally handicapped humans will fail to count as persons. If the capacity to communicate is a necessary condition, then Jean-Dominique Bauby, of *The Diving Bell and the Butterfly* fame, would have failed to count as a person after his catastrophic stroke—had he not worked out how to signal with his eyelid.<sup>vii</sup> (Can whether Bauby is still a person after his stroke really depend on contingencies about whether he can still communicate?) If persons necessarily have “the concept of a self as a continuing subject”, and believe that they are such selves, then David Hume, Derek Parfit, and most Buddhists are not persons.<sup>viii</sup> If you cannot be a person unless you are capable of valuing your own life, then you cannot be both a person and a committed nihilist about value (i.e. someone who finds himself compelled to believe that no such thing as value exists). If persons have to possess “emotionality”, then Mr Spock in *Star Trek* is not a person, and neither, perhaps, are some extreme autistics.<sup>ix</sup> And so on. Criterialists can make personhood sound rather exclusive; it can seem as hard to qualify for personhood as it is to make membership of the Country Club.

Of course, these problems are consequences of taking these “markers of personhood” as indicators of necessary conditions for personhood. A less extreme version of the criterial view might take each of the markers on some list, or some or other combination of the markers, as a sufficient condition of personhood: any creature that displays just one of these properties, or some large-enough variety of them, will count as a person. We might then want to argue about what counts as displaying a property, or possessing a capacity. Must I actually evince emotions to pass Warren’s emotionality test? How often, and how convincingly? Do I possess the capacity to communicate only when I have learned a language? Or just when I have learned to sign, or to get others to read my thoughts and feelings? Or do I have the capacity to communicate all along, just in virtue of being a member of a species that communicates, linguistically and in other ways?

Here we touch on familiar and long-running debates about potentiality versus actuality in our assessments of personhood. It is commonplace for criterialists of Singer’s sort to point out that potential prime ministers obviously do not have the rights of actual prime ministers, so that “potential persons”<sup>x</sup>, or those with the

potential for what Don Marquis famously called “a future like ours”<sup>xi</sup>, can hardly be treated as the same moral category as actual persons, or “us”. It is equally commonplace for the opponents of Singer-style criterialism to object that there must be something wrong with an emphasis on *actual* properties that makes it a live question whether one is a person while asleep or under general anaesthetic.

Clearly there are problems about at least some of the suggestions listed above about what the criteria of personhood might actually be. But not all those suggestions are implausible, and it would be easy enough, in principle, to weed out or refine the implausible ones. True, it is striking how little criterialists feel they need to do this; in practice, criterialists often seem perfectly happy to stick with what, intuitively, seem implausibly over-demanding criteria for personhood.

However, the plausibility or otherwise of the various suggested criteria is not my main concern here. Obviously enough, other criteria could be devised which did not have these consequences. The key difficulties for criterialism that I want to emphasise in this paper are three.

The first, which I can state very quickly, is about the argumentative method underlying criterialism. To put it at its simplest, it *looks* as if criterialists typically adopt their preferred criterion of moral significance because they think it explains moral intuitions which we have anyway. The criterion is adopted, in other words, because it doesn't have counter-intuitive consequences.<sup>xii</sup> But then, the same criterion is deployed, by the very same authors—Peter Singer and John Harris, for example—precisely to *produce* counter-intuitive consequences, such as the moral acceptability of killing small babies or the mentally disabled. There is, I think, a serious question as to what can possibly be going on here. It is not entirely obvious why building a moral theory out of one intuition or set of intuitions that we certainly have, and then using this theory to attack other intuitions that we also certainly have, is anything more than arbitrary selectiveness. Perhaps this is a question that all system-building moral theory faces.

The second difficulty for criterialism, which I shall also mention only very briefly, is that criterialism can seem to imply a quite implausible picture of moral reasoning. Are we really to suppose that what we do when we come across any being we haven't met before is *first* see whether it satisfies our criterion for personhood, and *then*, if it does, start treating it as a person? One is reminded of that well-known caricature the Cartesian Detective: we start with the hypothesis—the *opinion*, as Wittgenstein says in my epigraph—that the new being is a person, and feel justified in acting in line with that hypothesis, just insofar as the behavioural evidence confirms it. If criterialism is to be at all plausible, it cannot in this way be a general view about our typical real-time interaction with others. Such a picture might apply to some very exceptional cases, some of which we will discuss in section 4b. But in general, criterialism will have to be a view about the *criterion of rightness* for taking other creatures to be persons, not about *our decision procedure* for assuming that they are persons. For quite obviously, our normal interaction with others typically involves no such Cartesian-Detective rigmarole, and it takes a good deal of philosophical callousing of the soul not to find something absurd, and morally repugnant too, about the idea that it even could. But if in general criterialism only states the criterion of rightness, and not our decision procedure, for taking other creatures to be persons,

then it still faces a further objection, my third—which it is the main purpose of this paper to develop.

This third objection to criterialism, on which I shall spend the rest of this paper, is connected to this second point. My claim is that the sort of properties that criterialists home in on are not criteria of personhood at all. Rather, they are dimensions of interpretation of beings that we *already* take to be persons.

For my own part, I have always been aware of at least a vague sense of discomfort and unease here, the moment we start talking about these (or any other) properties as if they were criteria for personhood of the sort the criterialists have in mind. There is a hard-to-articulate feeling that, despite the considerable first-blush plausibility of criterialism, it leaves out something crucial; a feeling of the kind that is naturally expressed in ordinary conversation by a trailing-off “Yes, *but...*”. I suspect that my unease is not unrelated to the unease that Raimond Gaita is bringing out when he writes (Gaita 1991: 115) that it is

naïve to think that the kinds of properties which interest philosophers when they ask what is a person, and which we often share to some degree with higher animals, play the kind of role in our treatment of one another which is assumed by those who argue that differences in treatment must be justified by relevantly different properties of those kinds. That philosophical perspective from which we are encouraged to reassess our sense of how we might justifiably treat animals distorts and indeed cheapens our understanding of human life.

Gaita 1991 is a marvellous exploration of one philosopher’s attempt to make articulate that trailing-off “Yes, *but...*”. This essay offers another philosopher’s. My attempt is not necessarily the same as Gaita’s, but not necessarily in competition with it either.

And it is perhaps worth saying before I begin that what I am trying to do here, what Gaita was trying to do, is *hard*: hard to articulate, hard to state clearly, hard to see in the round, hard to apply. Again and again criterialism wins out in philosophers’ discussions of personhood because, conversely, it is dead easy; it is such a simple and straightforward view of personhood. There are plenty of spurious analogies between science and ethics around, but at any rate we can trust this one: no more in the case of the nature of the person than in the case of particle physics does the fact that a view is *simple* in any way improve its chances of being *true*.

## **2. The proleptic view of personhood**

Contrary to what criterialism seems to suggest, we do not look for sentience or rationality or self-awareness in a creature as a test to decide whether or not that creature counts as a person. It’s the other way round. Having once decided, on other grounds, that a creature *is* a person, we know that this makes it the kind of creature which is likely to display sentience, rationality, self-awareness and the rest of the personal properties. Hence we look for displays of these properties from the creature. That is to say, we treat it as a person *in advance* of any such displays.<sup>xiii</sup>

The biggest and clearest example of this is an activity that is absolutely central to human life: parenting. Parents are, of course, aware of the differences in rationality, linguistic capacity, self-representing ability, and so forth between young human beings at various ages and developmental stages. Nonetheless, a parent's attitude towards her child is always, basically, what Wittgenstein famously calls "an attitude towards a soul". She does not do what criterialism (especially if understood in decision-procedure terms) might seem to imply she should do: start out by treating her child as an inanimate object, like a sofa or a fridge or a rubber-plant, and grudgingly consent to adjust her attitude to it, one little step at a time, only as and when it proves itself more than inanimate by passing a succession of behavioural tests for at least having interests like a snail's perhaps,<sup>xiv</sup> and then later on for the interests of some more advanced kind of animal such as a puppy, and eventually for full criterialist personhood.<sup>xv</sup> If parents *did* treat their children in this almost behaviouristic fashion, the parents would be callous monsters, and the children would be basket-cases. In a world where parents generally accepted this sort of criterial view, and applied it directly in their parenting practice, even the best-adjusted of us would be a wolf-child.

Fortunately, however, parents are not criterialists, either in their decision procedure or in the criterion of rightness for ascriptions of personhood that they actually employ. Rather, a parent treats her child from the very beginning—and from before it is literally and actually true—as a creature that can reason, respond, reflect, feel, laugh, think about itself as a person, think about others as persons too, and do everything else that persons characteristically do. From the beginning her attitudes towards the child are not only "objective" (to use Strawson 1962: 67's slightly unfortunate term<sup>xvi</sup>): they also include what he famously calls "participant reactive attitudes", of just the same sort as she adopts towards anyone else.

Of course, as Strawson implies<sup>xvii</sup>, these participant reactive attitudes' scope may be more restricted in the case of a very small baby. The point is that they are there; and that like the baby—often indeed in advance of the baby—they *grow*. As Alasdair MacIntyre notes in a recent discussion of the development of practical reasoning in the individual person (1999: 90), ethically and psychologically adequate parenthood centrally involves "a systematic *refusal* to treat the child in a way that is proportionate to its qualities and aptitudes" (that is, its actual and non-idealised qualities and aptitudes, the kind of qualities and aptitudes that interest the criterialist).

The parent who says to her three day old baby, "What do you think? Should we give you some more milk now?", or to her three year old toddler "How kind of you to share your toys with your sister when she's sad!", is not deluded about a neonate's capacity for interpersonal practical reasoning, or a toddler's capacity for calculatedly empathetic altruism. But she is not making a sentimental joke either. She is treating the baby or the toddler proleptically, in the light of the ideal of personhood. By years of treating her children as creatures who "have the personal properties"—in the sense that interests the criterialist—she makes it true that they *are* creatures who have the personal properties in just that sense.

The child staggers across and plonks a book down on his sibling's lap; the parent's reaction is "How kind of you to let her share your book!" Is that, in fact, what the toddler was doing? The right answer to this question can be: "Yes it is—once the

parent has given this reaction.” The parent’s reaction is an interpretation of the child’s deed. Before the reaction, perhaps, there was no fact of the matter about what the little boy was doing: the child did not know himself what his action was. For all he knew about it, his action might as well have been simply a random sequence of bodily movements. But now that his mother offers her interpretation, the child, on his mother’s authority, learns to see himself a certain way. Because his mother frames his act as one with a certain meaning (as the giving of a gift), and because the child sees his mother as seeing his act this way, and because his mother sees the child as seeing her as seeing his act this way... (Grice 1957), the act *comes to have* that meaning.

Paul Grice’s famous work on meaning and intention is even more relevant to another clear example of this sort of prolepsis, namely language-learning. A baby begins by babbling—that is, by producing all the sounds the human mouthparts can make. Some of these sounds get a response because they are sounds that occur in the parents’ language, while others don’t because they are not. (The inability of many English adults to pronounce the Scots *ch* in *loch* or the Welsh *ll* in *Llanfair* is a learnt inability, not mirrored in English babies, who can and do make both sounds until conditioned not to by English adults.) The baby “homes in” on the sounds, and then on the patterns or combinations of sounds, that get a response. Then it learns to correlate particular patterns of sounds with particular contexts by the same homing-in process. What the parents say is “She’s learned to say ‘tiger’” (my own first child’s first word). What has happened is that the parents, by treating their daughter *as if she has* the word for, and then the concept of, a tiger, have made it true that she *has* that word. (And, a little later perhaps, that concept too; though I am not suggesting that any one word or concept can be learned on its own: in Wittgenstein’s familiar image, “light dawns gradually over the whole” (*On Certainty* 141).)

Stephen Darwall too talks of prolepsis when describing a third connected process, namely the process whereby we come to be, and to hold others, morally responsible (Darwall 2006: 87-88):

There seem to be many cases where we wish to hold others accountable though we seem to have very good evidence that they are not free to act on moral reasons in the way our practices of holding someone fully responsible seem to presuppose... In some instances, for example with children, we seem simultaneously to move on two tracks in the process of inducting them into full second-person responsibility, sometimes treating them proleptically as though they were apt for second-personal address as a way of developing moral competence while nonetheless realising... that [at their developmental stage] this is an illusion.

Or an illusion, at any rate, if you agree with the criterialist about what it takes to be “apt for second-personal address”, and/or about what it takes to “have the personal properties”. My point, with which Darwall might or might not agree—I’m not sure—is that what it takes does not have to be already-realised and –actualised capacities of the sorts that interest criterialists. (The mother who asks her neonate child “Shall I give you some milk?” is certainly engaging in second-person address, *despite* the manifold incapacities, perfectly well-known to her, of neonates.)

There is at least this much truth in Dennett 1987's well-known idea of the "intentional stance": persons constitute each other as persons, both as agents and as patients, by *treating* each other as persons, as agents and patients. Something like the Davidsonian "principle of charity" (Davidson 1980) is at work in our mutual interpretations. By charitably, and proleptically, interpreting the other as a person, I *make* him a person.

The looseness and ambiguity of the term "person" is evident when we say things like this. In ordinary language, "person" can certainly, at one end of its semantic range, mean something like what Singer and other criterialists want it to mean, a "rational and self-conscious being" (or whatever). Yet at the other end of its semantic range, the ordinary-language term can also be a synonym for "human animal". It seems to be part of the *point* of our person-concept to allow those young humans who are persons only in the "human animal" sense to enter into a continuum of meaning towards the far end of which they can become persons in much stronger senses.

Many things in our shared life can be seen in this sort of way as aspirations towards an ideal which, like any ideal, is never fully realised. The basic sort of idealisation, the sort that grounds attributions of personhood, is not the only sort. We ascribe the rights to freedom of assembly and freedom of speech to every human being, even though we know that small babies can't speak and don't assemble (not without help), and that acutely asocial loners may well never (or at least never again) actually exercise either right. The behaviour of parents, besides (as described) involving the person-constituting prolepsis in a particularly clear form, is also full of other similarly-patterned proleptic idealisations that might, on a criterialist's view, be regarded as quite unrealistic. Parents give their children the opportunities to play the cello, to learn Spanish, to use their monthly allowances sensibly, to be polite to Great Aunt Maud (...), even though the parents may be fully aware that every one of these opportunities will in all probability be spurned. In minimally decent jurisdictions the treatment even of hardened criminals displays the same pattern of basic *openness* to the convict, however much of an unrealistic idealisation this openness may seem to involve: the jurisdiction goes on offering the criminals opportunities to reform, even though it is as good as certain that these opportunities will not be taken up. (I think that something like this refusal to "close the door" finally and irrevocably on the criminal is also, in connection perhaps with some thoughts about the state's obligation to be fundamentally benevolent to the citizens for whose sake alone it after all exists, the ground of a good reason to reject capital punishment. But that's another story.)

In general, the idealisation that accords X the status due to a  $\phi$ er does not have to involve any kind of expectation that X will ever actually  $\phi$ , or even can  $\phi$ . What it *does* have to involve is the understanding that X is of a kind that characteristically includes  $\phi$ ers. The human kind includes individuals who characteristically (at the right point in the human life-cycle) will freely assemble and speak. That is why it is not ridiculous to give the rights to free assembly, speech, and political action, e.g. becoming a senator, to any and every human, including asocial misanthropes, those (as yet, or as of recently) physically incapable of speech, and those who choose never to utter a word. Contrast the equine kind, which does not characteristically include any such individuals. That is why it *is* ridiculous to give these rights (or the status of a senator) to any horse.



Considered as a kind, children characteristically include individuals who learn the cello or Spanish or wise use of money or good manners to venerable relatives (well... eventually). That is why the parents who try to teach these lessons to their children are not acting absurdly even if they know in advance that *this particular* child is not going to respond—whereas parents who tried to teach these lessons to the family goldfish or the rubber plant *would* be acting absurdly. Similarly, there is nothing absurd about offering the old lag a place on the rehabilitation programme, even if it seems certain that *this* old lag will always refuse or abuse the offer: and the reason why not has to do with the general fact that some old lags do sometimes reform.

*A fortiori*, there is nothing absurd about giving the rights of free speech, assembly, private choice and the rest of it to women, or to other races, or to homosexuals or transgender people, or to the disabled, or to gingers, or to the various other sorts of historically marginalised groups. People in these categories are not another kind, as horses and hamsters are. Their well-being is *human* well-being, just as white male heterosexuals' is, and just as horses' and hamsters' well-being is not. That, in a nutshell, is why speciesism is *not* like racism or sexism.

Clearly, then, this is not just a point about the induction of babies and very small children into the social life of humanity at large. All humans, even the healthy intelligent mature well-adjusted independent affluent adults that typical criterialism is so clearly focused on—the Country Club again—are always only incompletely and impurely agents. Incompletely, because we are never all agents can be. However well things go, we never do or even can reach the full potential of human well-being, which would include a fully worked-out articulation of our own agency, a full coordination and marshalling of our own forces and powers for action in the world: towards this we are always only, at best, on the way. And impurely, because each of us always *can* be interpreted, by a sufficiently determined sceptic, as not really an agent at all, but a victim of the instinctive or other determining external forces that constantly threaten anyone's agency. To extend “an attitude towards a soul” towards *anyone* is always, and not just with children, an idealisation. As Tamar Schapiro notes within a Kantian framework (Schapiro 1999: 723):

[T]here is a sense in which no one, regardless of age or maturity, is able to achieve autonomy on Kant's view. This is because the notion of autonomy in Kant is an ideal concept which outstrips all possible realisations in experience. Strictly speaking, every instance of human willing is necessarily an imperfect realisation of transcendental freedom, and every virtuous character necessarily falls short of perfect virtue. And yet the applicability of the moral law depends on our mapping these ideal concepts onto ourselves and one another for the purposes of guiding action. So we are to regard the social world as a community of autonomous agents despite the fact that perfect realisations of autonomy are nowhere to be found.

The criterialists are right to insist that properties like self-awareness, emotionality, rationality, and so forth are crucial parts of our concept of a person. It does not follow that these properties can be used as the criterialists want to use them: as *tests* for personhood. Even at the cool-hour level of the criterion of rightness—never mind the

real-time level of deliberative procedure—to treat someone as a person is not to put a tick in the box by her name, to show that she has passed some inspection or met some standard, of rationality or self-awareness or emotionality or whatever. Indeed, it seems no less arbitrary discrimination to say “Sorry, you’re not rational or self-aware or linguistically capable or emotionally responsive enough to count as a person” than to say “Sorry, you’re not white enough to count as a person”. Behavioural properties like rationality, self-awareness, emotionality are not tests for, but parts of the ideal of, personhood. To treat someone as a person is to engage with him as the kind of creature to which that ideal applies. So to treat him is not, at the deepest level, a response to his *behaviour* at all, but to his nature. To see some creature as a person is to take an attitude to that creature which, *before* any behavioural evidence comes in, is already different from our attitudes to creatures that (we think) aren’t persons. This is the second-person standpoint (to use the title of Darwall’s outstanding book): the attitude we take to persons, and do not take to sofas, fridges, or rubber plants.

The fact that it takes charity, in more than one sense, to constitute persons imposes on persons a crucial sort of vulnerability and dependence: “The personal reactive attitudes rest on, and reflect, an expectation of, and demand for, the manifestation of a certain degree of good will or regard on the part of other human beings towards ourselves” (Strawson 1962: 70). All human persons are, in Alasdair MacIntyre’s splendidly apt phrase, “dependent rational animals”, as MacIntyre himself emphasises (MacIntyre 1999: 81-85):

In most moral philosophy the starting point is one that already presupposes the existence of mature independent practical reasoners whose social relationships are the relationships of the adult world. Childhood, if noticed at all, is a topic that receives only brief and incidental attention... To become an effective independent practical reasoner is an achievement, but it is always one to which others have made an essential contribution. The earliest of these relate directly to our animal existence, to what we share in our development with members of other intelligent species... [But what] we [also] need from others, if we are not only to exercise our initial animal capacities, but also to develop the capacities of independent practical reasoners, are those relationships necessary for fostering the ability to evaluate, modify, or reject our own practical judgements... Acknowledgement of dependence is the key to independence.

We are vulnerable to each other, and dependent on each other, because it is remarkably easy to attack persons by withdrawing the second-personal attitude from them. Part of the reason why it is so easy to stop seeing others as persons is, as I said before, because the interpersonal attitude always involves a degree of idealisation. There is always *some* truth in sceptical or cynical views that reject that idealisation. (For example, the human person *is* a physical object—even if it is not “merely” a physical object.)

We fail to see others as persons in a small way when we explain away what others do, or refuse to listen to them or to take their wishes and ambitions seriously. (Parents are never perfect in charity; children’s protests against such treatment are a recurring theme of family life.) We do it in bigger ways when we “use” others, or (in Kant’s phrase) fail to treat them as “ends in themselves”. At the limit, we can withdraw the interpersonal attitude from others completely, by pretending that they

simply aren't there, as British and Irish settlers did to the aboriginal Australians under their notorious legal doctrine of *terra nullius* (see Gaita 1998). Or we can do it by treating them merely as physical objects, as often happens in pornography, violence, and murder. Or, lastly, we can do it by declaring them “non-persons”, as the Nazis did to the Jews—and as criterialists routinely do to the very old, the very young, and the very disabled.

The personal properties, then, set the content of the ideal of personhood. They do not set the boundaries of the class of persons (the PMC, as I've also called it). What does set the boundaries? That is my question in section 3. My answer will be the humanist answer already advertised in section 1: roughly, the boundaries of the class of persons are the boundaries of the human species.

### 3. The truth in speciesism

Rubber plants, fridges, sofas, human babies: we come to the last of these four sorts of things with a very different attitude from our attitudes to the other three sorts. And my point is that we *come to them* with different attitudes. Our attitudes to them are not responses to evidence that we gather *after* we have “come to them”. *The moment I meet* a rubber plant or a fridge, I am perfectly sure, unless my sanity is in question, that it will not attempt to talk to me or tickle me—as, say, a baby might.

What makes me so sure of that? Well, simply the sort of thing a rubber plant or a fridge is, and the kinds of behaviour that it is natural to expect from that sort of thing. Rubber plants and fridges are not the kind of thing that tries to talk to you or tickle you. Babies are. And if you ask “What kind of thing is that?”, the obvious answer is: “Young human beings”. If some creature is a human being, it is completely natural to expect it to be the kind of thing that might (at least if it's a family member) try to talk to you, or tickle you, or otherwise demonstrate its (budding) mastery of the kinds of mental capacity that criterialists take to be criteria. (Talking to someone and tickling someone both require interesting interpersonal capabilities. To do either, you have to want things to be a certain way for someone else—roughly, you have to want to get them to think a thought, or to feel a tickle. These are wants with quite sophisticated contents.)

This sort of expectation is based squarely on the nature of the creature in question. And that means, on its nature *as a member of the human species*. Our understanding of which creatures are persons, and so may be expected to exhibit the personal properties, is not formed on an individual basis but on the basis of generalisation from experience of humans in general. I have already mentioned Wittgenstein's remark that “my attitude towards him is an attitude towards the soul”; on the same page of *Philosophical Investigations*, Wittgenstein also says that “the human body is the best picture of the human soul” (*der menschliche Körper ist das beste Bild der menschlichen Seele*). And perhaps what he means is this: that when we see that some creature is a human, just seeing that is enough to license us to adopt the interpersonal attitude to that creature.

We know in advance how we can expect any particular rubber plant or fridge to behave, because we can draw on a bank of past experience (our own and others') of

rubber plants and fridges. In this sense, the fact that a thing is a rubber plant or a fridge sets the scope of our expectations about it. In just the same way, we know what kinds of capacities and properties humans in general have. It is that background knowledge that we draw on, whenever we apply the interpersonal attitude to any individual human being, simply on the basis that it *is* a human being.

We obviously do the same with other species of animals. When I meet a cat or a gerbil, I treat it neither as an inanimate object like a sofa or a fridge or a rubber plant, nor as a person. (Not, of course, that I treat all inanimate objects the same, either. Sofas, fridges, and rubber plants are very different things, especially rubber plants.) You can play games with a gerbil. With a dog or a cat, you can even have a cuddle (in the reciprocal sense, I mean).<sup>xviii</sup> Neither is possible with a fridge. On the other hand, you would be wrong to expect a dog to play a good hand of contract bridge, or a cat to give you good investment advice, or a gerbil to read your mind, as a human being might. What you expect of an individual dog or individual cat will be different from what you expect of an individual gerbil, and both will be different from your expectations of inanimate objects. In each case, your expectations will be set by your background knowledge of the different kinds in question. Where dogs, gerbils, cats, and indeed all other kinds of living thing are concerned, this obviously means the different *species* in question. Why should it be otherwise with humans?

These remarks should enable us to see—*pace* the criterialists—how central it is, to our moral responses to the things and creatures around us, to assign them to species. It is species-assignments that enable us to answer what you might reasonably think is the most basic question of all about any thing that might be in front of us at any time: namely, what kind of thing it is. Until we know what species it is, we do not know how it may act, or how to treat it: what is good for it, how we might harm or hinder it, what counts as health for it, and so on. The nature of the creature determines what well-being is for it; and its nature is its species. To put it another way, the best way to answer “What is it?”, when this question is asked about any individual animal, is to give the name of its species. Or as we might ask moral individualists like McMahan: “If species-membership isn’t an intrinsic property, what *is*?” Singer, Rachels, and McMahan never tell us why *being human* cannot count as any creature’s “own particular characteristic”. It can’t be merely because *being human* is a group membership (as McMahan 2005 suggests). For there is nothing to stop a given property from being both a “group membership” and an “own particular characteristic”; that is how it is with *being a vertebrate*, for instance. To put it more formally, nothing stops a given property from being both a relational property, and also an intrinsic property.<sup>xix</sup> Indeed, if Kripke is right in his well-known thesis of the metaphysical necessity of origin, such properties must be very common.

This shows how humanism can be true without falling prey to a charge of arbitrary speciesism. To treat any species of creatures as persons is to treat them, not as successful passers of some test for the personal qualities, but on the assumption that they are individuals of a species the ideal for which is (roughly) the highest attainable development of those qualities. In this sense, references to species are an ineliminable part of our talk about persons.

In that sense, but not in some others. We are now in a position to respond to four familiar objections to “speciesism”. I consider these in section 4. The aim of my

response is not of course to defend speciesism, if that is defined as “*arbitrary* moral discrimination on the grounds of species”. But it is to defend humanism, the view that not all moral discrimination on the grounds of species *is* arbitrary.

#### 4. Four objections to speciesism

*a. All animals are equal.* Peter Singer (1993: 57) writes that “No matter what the nature of the being, the principle of equality requires that the suffering be counted equally with the like suffering of any other being”. Nothing I have said so far commits me to rejecting this application of the principle of equality. If I were to accept it, I would want to add, first, that plenty of other things besides suffering “count”, and secondly, that however alike human and animal *pain* (the physical sensation) may be, human and animal *suffering* (the state of mind) are mostly quite unlike. A cat can be tortured like Gloucester, but it cannot suffer like King Lear.<sup>xx</sup> And suffering matters more than pain.

But perhaps we should reject this application of equality anyway. Right now there is (quite likely) monstrous pain, and a fair amount of suffering too, on the plains of the Serengeti, where a wounded buffalo is being eaten alive by a pride of lions. Singer’s principle of equality requires us to say that our duty to prevent the buffalo’s pain and suffering in the teeth of the lions is directly proportional to the amount of pain and suffering involved. If any implication of any moral theory is absurd, this one is. There just isn’t any duty, not even a *prima facie* one, for us to act as Zoological Pain Police. (Don’t say: “We can’t intervene, because the lions might eat *us*”. Where is your moral courage? And don’t say: “But it causes less pain in nature overall if we don’t intervene in nature in this sort of way.” As many writers on the theological problem of evil are keen to remind us, most animal suffering is gratuitous. If lions can be caught and trained to jump through hoops, perhaps they can be caught and trained to kill their prey cleanly before eating it. So is anyone *trying* to train them? Of course they’re not.) The vast majority of the huge amount of animal pain and suffering that happens every day is simply none of our business. “It is not an accident or a limitation or a prejudice that we cannot care equally about all the suffering in the world: it is a condition of our existence and our sanity” (Williams 2006: 147).<sup>xxi</sup>

So it looks like Singer’s principle that all suffering counts for the same, no matter what the species involved, is just false. On the other hand, some weaker principle of this general sort seems clearly true: perhaps “Animal suffering is bad for the same sort of reasons as human suffering is bad”, or something like that. What also seems clearly true is that humans should do what they can not to generate animal suffering, and to prevent animals from being needlessly hurt or damaged *by humans* (and sometimes, though less often, by accident or by each other). This is a much weaker principle than Singer’s equality-of-suffering principle. But perhaps even this is strong enough to generate a good argument for vegetarianism. I suspect it all depends what counts as needless hurt or harm.

Hence “the limit of sentience” is not “the only defensible boundary of concern for the interests of others” (Singer 1993: 58). Certainly that limit marks one distinction between the ways we should take an interest in the welfare of other creatures. For example, hacking a frog to pieces is wrong in a way that hacking an

orange to pieces is not, partly because an orange certainly cannot feel pain, and a frog pretty definitely can. But only partly. Quite aside from the pain involved, destroying a living creature for fun is wrong in a way that destroying a piece of fruit for fun is not. (Hacking trees, or anaesthetised frogs, to pieces for fun is usually wrong too.<sup>xxii</sup>) Unless we think that nothing matters except pain and pleasure—and this is not, e.g., Singer’s position either (Singer 1993: xi)—there is no reason to think that the distinction between sentience and non-sentience is the only fundamental moral distinction we can make among beings; nor that it is the most important.

There are some interesting complications about the notion of sentience. If “sentience” means simply “capacity to feel (sensational) pain and pleasure”, then sentience does not fit the story that I have developed about how the personal properties come into being. The capacity to feel pain- and pleasure-sensations is certainly one of the capacities that we expect to find in any human person. But it is not, unlike rationality or intentionality, a capacity whose presence in any person is even partly constituted by others’ idealising assumption, expectation, or hope that it is there. Whether or not a being can feel pain does not depend on others’ adopting “participant attitudes” or taking the “intentional stance” towards him, in the way that (I have argued) his capacities for self-consciousness, rationality, and suffering do. I don’t think this shows that sentience is a counter-example to my thesis about how the person-constituting properties are developed. Rather, I think it shows that sentience is not one of the person-constituting properties. A creature which had the (other) person-constituting properties, but lacked the capacity for pain- and pleasure-sensations, would still undeniably be a person.<sup>xxiii</sup> A creature which had the capacity for pain and pleasure, but lacked the (other) person-constituting properties, surely would not be.

*b. Aliens, talking chimps, Locke’s parrot, and a possessed fridge.* The position that I am defending under the name of humanism says that, normally, to be a person it is *sufficient* to be a human being. Not *necessary*, because humanism does not imply that members of other species could not be drawn into the moral community of human persons. On the contrary, I can see at least three distinct ways in which this could happen. One is that we could discover a whole species—aliens, or angels, or what have you—which is like the human species in that its members are characteristically persons (in one sense or another). A second is that a species already known to us might, *en masse*, somehow change so as to become characteristically persons. Or, third, some particular individual creature—Locke’s parrot, say (Locke, *Essay* 2.27), or Washoe the chimp (Singer 1993: 111)<sup>xxiv</sup>, or some group of creatures, like Peter Carruthers’ Chimp Class of ’92<sup>xxv</sup>—might start to count as persons, perhaps e.g. by displaying the personal properties, even though this behaviour is otherwise completely untypical of the species in question. (Come to that, I suppose a possessed fridge might, in principle, start displaying the personal properties.)

In these special cases, we perhaps *will* have to appeal to something like the criterial view of personhood as our decision procedure for the vexed cases. Apparently we will have to ask whether this alien species, or this new class of highly-educated chimps, or this individual chimp, parrot, or spooky fridge, displays something like the criterialist’s personal properties, *before* we can decide whether to count these beings as persons.

Doesn't this show that my arguments against criterialism were mistaken? Not at all. It does show that something like criterialism may perhaps be our decision procedure if and when we encounter a being *that we do not already know how to classify as a person or not a person by way of its species*. (Perhaps; actually I doubt even this much is true, but at any rate it is the most that is true.) However, it would be utterly misleading to generalise from thought-experiments about these special and rare cases—almost all of which, to date, are imaginary—to alleged conclusions about the normal cases. (It would be as misleading to do that as it would be to generalise the justificatory/ rational structure of *every* belief we have about the external world from the justificatory/ rational structure of the external-world beliefs that we might form when we are thinking about the classic, and nearly always imaginary, sceptical predicaments.) Unfortunately, just this generalisation is a criterialist stock-in-trade. The most famous example of it is perhaps Mary Anne Warren's derivation of the personal properties (her list was quoted in section 1) from an imagined meeting with aliens. But even if Warren is right that this is the decision procedure that we would have to adopt with *aliens* to settle whether they were persons in some extraordinary imaginary scenario, nothing at all follows about our beliefs (either at the decision-procedure or at the criterion-of-rightness level) about whether any *humans* are persons in a multitude of ordinary real scenarios. For we have met humans before.

But anyway, I suspect Warren's description is much too black-and-white to get anywhere near characterising the decision procedure that would really be involved in such a meeting. (This is why I say, above, only that we would probably have to use *something like* the criterialist approach in alien encounters and similar.) For one thing, there is not the slightest reason to think that, in such a meeting, *yes* and *no* would be the only possible answers available to the question "Are these persons?". Even *Star Trek* yields a rather subtler answer—"It's life, Jim, but not as we know it"; the world, after all, is a complicated place, and it constantly throws us surprises. For another, our responses in such situations would be—or should I say *will* be?—based on all sorts of factors and reactions, most of them probably defying explicit articulation; on judgements about "forms of life", in fact. It would not and could not be based on some neatly finite check-list of diagnostic properties like Warren's. Looking beyond *Star Trek*, science fiction is full of wonderfully complex and subtle explorations of such encounters. The contrast with the mechanical simplicities of a criterialist list such as Warren suggests is instructive—and not *entirely* to be dismissed as unfair on the grounds that Warren is, after all, not writing science fiction.

*c. The two-species scenario.* A third interesting objection to speciesism is offered by Peter Carruthers:

Suppose it had been discovered that human beings in fact consist of two distinct species, otherwise hardly distinguishable from one another, the members of which cannot inter-breed. In these circumstances it would plainly be objectionable for the members of the majority species to attempt to withhold moral rights from the members of the minority, on the mere ground of difference of species. This, too, would be obvious speciesism. (Carruthers 1992 Ch 3)

Carruthers, I think, is quite right to claim that it would alter nothing morally speaking if we discovered that "the human species" was in fact two different

biological species—perhaps *homo neanderthalis* and *homo sapiens*—living together in a single moral community. (Modulo the point about not interbreeding, some anthropologists think that this has actually happened: Trinkaus and Shipman 1993.) Pace Carruthers, his argument does not show that there is no species-classification that bestows participation in the moral community on its members, and so that species-classifications are morally irrelevant. What his argument shows is that there may be *more than one* species-classification (or other kind-classification) that bestows participation in the primary moral constituency on its members. That is something that we should grant anyway given the possibility of other species whose members are characteristically persons; and it is not inconsistent with the claim that species-classifications are morally significant.<sup>xxvi</sup>

*d. Species and Darwinism.* A fourth objection to my argument is the claim that species don't exist: that as a matter of strict biology, there is no clear differentiation between humankind and "other species", only a differentiation between different parts of the same evolutionary family tree. Species therefore are populations, not Aristotelian (or Kripkean) natural kinds (Sober 1994). True, but irrelevant to my argument, which can work equally well with the population conception of species, or indeed with any conception at all that will allow "humanity" to count as the name of a kind of *some* kind. At least for creatures like us, whose temporal experience runs only over decades, hardly ever over centuries even, let alone the millions of years that speciation takes, it should not be a controversial thought that a population of creatures sharing a common genetic heritage, physiology, and ethology can sensibly be treated as a unitary grouping, as in fact common sense treats it. And that thought is all I am committed to meaning by "species".

So it looks like humanism can resist these four objections. And it looks like humanism can accept, indeed welcome, the thought that humans are not the only persons, the only members of the primary moral constituency—or at least, they might easily not be. What about the question on the other side: the question whether all human beings are persons? I turn to this question in the next and final section.

## **5. Are all human beings persons?**

Humanism as I have developed it here straightforwardly implies that all human beings are persons. Our treatment of any human being should be conditioned by the background of expectations, hopes, and aspirations that spell out what we know, from experience, humans in general can be. *Eudaimonia* in its broad outlines is the same for all human beings; and requires, as a general rule, that we must give all human beings the space to achieve *eudaimonia*—whether or not they predictably *will* achieve *eudaimonia*. To deny this space to any individual human being is to exclude that individual from the moral community of persons. And that is a serious injustice.<sup>xxvii</sup>

The very young are not excluded from the moral community of persons, just because they have not yet achieved many key forms of human *eudaimonia*; the mentally handicapped are not excluded from the moral community of persons, just because they cannot ever achieve many key forms of human *eudaimonia*. As I pointed out in section 1, *all* persons are incomplete and impure in their agency. No one ever



does everything she could with her own nature (partly, of course, because there are just too many things she could do). In one way or another, *all* of us are less than fully free to pursue happiness; everyone is wounded or damaged or limited or compulsive, or just plain ill, in some respect. That does not make it wrong to treat “normal” humans, whatever their age or IQ, with an interpretive charity, and with a kind of pity, that sees beyond their limitations, and leaves the door open, in each individual’s case, to what *any human* might become—even if we are well aware that *this person here* has no chance of becoming that.

[T]he kind of care that was needed to make us what we have in fact become, independent practical reasoners, had to be, if it was to be effective, unconditional care for the human being as such, whatever the outcome. And this is the kind of care that we in turn now owe or will owe. Of the brain-damaged, of those almost incapable of movement, of the autistic, of all such we have to say: This could have been us. (MacIntyre 1999: 100)

There is the possibility of deep moral concern for retarded people, in which they are seen as having, however incomprehensible we may find it, a human fate, as much as anyone else’s.... Someone may be very touched by the response of a severely retarded person to music; and there may be in that being touched an imaginative sense of shared humanity. (Diamond 1991: 55)

The terminally ill are not an exception to this rule either, just because they cannot any longer achieve many key forms of human *eudaimonia*. Perhaps it is true of them that they no longer have—to reuse Marquis’ phrase—“a future like ours”, and so will not achieve very much, or any, more *eudaimonia* before they die. Given that their futures have closed in on them in this radical way, you might call it “futile” to go on treating them with the full range of interpersonal hopes and expectations that you extend to any human, or to go on being as scrupulous about not killing them as you would be about anyone else. There again, you might also call it “respectful”; you might call it an expression of *pity*.

To pity someone in the sense I have in mind is to think, as MacIntyre puts it, “this could have been us”: this person too is part of human life and shares in human ideals, even if the idea of realising them in almost any way is quite hopeless for him/her. Such pity is a key part of our charity towards other people; of what is involved in recognising them as *people*, as fellow human beings.

(It is interestingly not something that follows in any very clear way from an Aristotelian virtue ethics where the focus is, eudaimonistically, on *maximal* forms of flourishing—on how much you as a being can develop, not on how many relatively undeveloped beings you can extend charity towards. Indeed it is a notorious fact about Aristotle that there is something rather ruthlessly bright-eye-and-gleaming-coat<sup>xxviii</sup> about his conception of *eudaimonia*, and that—not unconnectedly—he seems to have had little room in his ethics for the notion of pity.<sup>xxix</sup> Pity is, however, part of the sorts of virtue ethics that we may find in a Christian writer like Simone Weil. And that is the sort that I would defend, though I cannot defend it here.)

There are, as many people have emphasised, better and worse ways to die. And one of the worse ways to die is to have recognition of your humanity withdrawn

from you before it happens. Part of understanding why that is such a bad thing to have happen to you must, I think, depend on taking a different view from the criterialist's about the locus of the value of persons. For the criterialist, apparently, the value of persons consists in their instantiating the person-constituting properties. Aside from such instantiations, persons, at the most basic level, *have* no value (compare the familiar view in bioethics that being alive in itself is neither good nor bad); indeed there is a sense in which, aside from such instantiations, there *are* no persons. What such a view of the value of persons obscures from view is the notion of loving an individual in itself; it seems to be only the *properties* of an individual that a consistent criterialist can love—and then, only when they are duly manifested. But the idea that the object of love is always and necessarily some property-instantiation, common though it is, is a dogma. If property-instantiations can be objects of love (we may fairly ask), why not things in other metaphysical categories: individual human beings, for example?<sup>xxx</sup>

Alongside this dogma, perhaps there is another, related but slightly different, dogma at work in securing the foundations of criterialism. This is the dogma that whenever we answer the question “Why is X valuable?” by citing some valuable property V that X displays, what we must be really saying is always that it is (this instance of) V that is valuable, not X. But this too is a dogma, and unmasking it may be another way of setting ourselves free from the lures of criterialism. To cite V in explanation of the claim that X is valuable can, so to speak, be understood the other way around: you can take it that citing V is displaying a *symptom*, a *reminder*, of the value of X. When George Orwell, in a famous passage from his essay “A Hanging”, describes the growing fingernails and the refusal to walk through a puddle of the man who is about to be hanged, I take it that this latter thing is what he is doing.<sup>xxxi</sup> He is not saying that having growing fingernails is a value-giving property. Rather he is saying that *being human* is a value-giving property, the importance of which can be brought out for us by all sorts of exercises in the “assembling of reminders”—including such reminders as the growth of fingernails.

“But all the same, isn't there a profound difference between what we hope for *in general* for other people, and what we hope in the particular case of someone who is severely mentally handicapped?” There certainly is. There is always a distinction between the aspirations for other persons that are grounded in human *eudaimonia* in general, and the aspirations for specific other persons that are grounded in our particular knowledge of them as individuals. This distinction is just as applicable to the profoundly mentally handicapped as it is to anyone else. And it makes for the same commonalities, and the same differences, as with anyone else. If your friend is tone-deaf, there is no need to pretend that he will ever be Mozart; but that does not make it all right to take home all his CDs. If your friend loses both his legs in a skiing accident, you needn't spend the rest of his life pretending to him that he can still ski; but you miss something vital, too, if you bounce into his hospital ward straight after his double amputation with the words “So, you won't be wanting those ski-boots any more”. Many of the details of what counts as just and loving treatment of other persons are, certainly, set by the particular trajectories of their particular lives. But the most basic and elementary requirements of love and justice, is set by reference to the aspirations that arise from human *eudaimonia* in general. And this applies just as much to the mentally disabled as to anyone else.

In Nick Hornby's very funny novel about suicidal depression, *A Long Way Down*, Maureen, the mother of a severely mentally disabled child, says this:

This all began years ago, when I decided to decorate his bedroom. He was eight, and he still slept in a nursery—clowns on the curtains, bunny rabbits on the frieze round the wall, all the things I'd chosen when I was waiting for him and I didn't know what he was. And it was all peeling away... and I hadn't done anything about it because it made me think too much about all the things that weren't happening to him, all the ways he wasn't growing up. What was I going to replace the bunny rabbits with? He was eight, so perhaps trains and rocket ships and footballers were the right sort of thing for him—but of course, he didn't know what any of those things were... But there again, he didn't know what the rabbits were either, or the clowns. So what was I supposed to do? Everything was pretending, wasn't it? The only thing I could do that wasn't make-believe was paint the walls white, get a plain pair of curtains. That would be a way of [saying] that I knew he was a vegetable, a cabbage, and I wasn't trying to hide it. But then, where does it stop? Does that mean you can never buy him a T-shirt with a word on it, or a picture, because he'll never read, and he can't make sense of pictures? And who knows whether he even gets anything out of colours, or patterns? And it goes without saying that talking to him is ridiculous, and smiling at him, and kissing him on the head. Everything I do is pretending, so why not pretend properly?

In the end, I went for trains on the curtains, and your man from *Star Wars* on the lampshade. And soon after that, I started buying comics every now and then, just to see what a lad of his age might be reading and thinking about. And we started watching the Saturday morning television together, so I learned a little bit about pop singers he might like... thinking about these things helped me to see Matty, in a strange sort of way... I made up a son. (Nick Hornby, *A Long Way Down*, pp.118-9)

Of course there is—as Maureen herself recognises—something delusional, and something pathetic, in her treatment of Matty. For sure, her attitude to him is strained and exaggerated. But what we should not lose sight of is that it is a strained and exaggerated version of something morally indispensable: the kind of interpretive charity, the humanistic understanding and pity, that, I have argued here, we all need from everybody else all the time. In the words of an older writer: “Charity bears all things, believes all things, hopes all things, endures all things.”<sup>xxxiii</sup>

## Bibliography

- Bauby 1998: Jean-Dominique Bauby, *The Diving Bell and the Butterfly*. London: Fourth Estate, 1998.
- Carruthers 1992: Peter Carruthers, *The Animals Issue: Moral Theory in Practice*. Cambridge: Cambridge University Press.
- Chappell 1998: Timothy Chappell, *Understanding Human Goods*. Edinburgh: Edinburgh University Press.
- Chappell 2003: Timothy Chappell, "Absolutes and particulars", 95-117 in *Modern moral philosophy*, ed. A. O'Hear. Cambridge: Cambridge University Press.
- Davidson 1980: Donald Davidson, *Inquiries into truth and interpretation*. Oxford: Clarendon.
- Darwall 2006: Stephen Darwall, *The Second-Person Standpoint*. Cambridge, MA: Harvard University Press.
- Diamond 1991: Cora Diamond, „The importance of being human“, in David Cockburn, ed., *Human Beings*. Cambridge: Cambridge University Press.
- Engelhardt 1989: Tristram Engelhardt, "Ethical Issues in Aiding the Death of Young Children", in *Euthanasia: the moral issues*, edd. Robert M Baird and Stuart E Rosenbaum. Buffalo, N.Y. : Prometheus Books.
- Gaita 1991: Raimond Gaita, *Good and Evil: an absolute conception*. London: Routledge.
- Gaita 1998: Raimond Gaita, *A Common Humanity*. London: Routledge.
- Grau 2010: Christopher Grau, "Love and history", *Southern Journal of Philosophy* **48** (2010), 246-271.
- Grice 1957: H.P.Grice, "Meaning", *The Philosophical Review*, **66**: 377-388.
- Harris 1985: John Harris, *The Value of Life*. London: Routledge.
- Alison Hills, "Utilitarianism, contractualism and demandingness", *Philosophical Quarterly* **60** (April 2010), 225-242.
- Nick Hornby, *A Long Way Down*. London: Penguin, 2006.
- David Hume: *A Treatise of Human Nature* [1739].
- Kahane 2010: Guy Kahane, "Feeling pain for the very first time: the normative knowledge argument", *Philosophy and Phenomenological Research* **80.1** (2010), 20-49.
- Long and Sedley 1987: A.A. Long and D. Sedley, *The Hellenistic Philosophers*, Cambridge University Press, Cambridge, Vol.I, 67A
- Marquis 1997: Don Marquis, "An argument that abortion is wrong", in Hugh LaFollette, *Ethics in Practice*, First Edition (Oxford: Blackwell, 1997).
- MacIntyre 1999: Alasdair MacIntyre, *Dependent Rational Animals: why human beings need the virtues*. Chicago: Open Court, 1999.
- McMahan 2005: Jeff McMahan, "Our fellow creatures", *Journal of Ethics* **9** (2005), 353-380.
- Olson 2007: Eric Olson, Stanford Encyclopedia of Philosophy entry on "personal identity", <http://plato.stanford.edu/entries/identity-personal/>
- Parfit 1984: Derek Parfit, *Reasons and Persons*. Oxford: Oxford University Press.
- Patterson and Gordon 1993: "The Case for the Personhood of Gorillas", in Paola Cavalieri and Peter Singer (eds.), *The Great Ape Project* (New York: St. Martin's 1993), 58-77.
- Rachels 1990: James Rachels, *Created From Animals: The Moral Implications of Darwinism*. Oxford: Oxford University Press.
- Schapiro 1999: Tamar Schapiro, "What is a child?", *Ethics* **109.4**: 715-738.

- Singer 1983: Peter Singer, *The Expanding Circle*. Oxford: Oxford University Press.
- Singer 1991: Peter Singer, *A Companion to Ethics*. Oxford: Blackwell.
- Singer 1993: Peter Singer, *Practical Ethics*. Cambridge: Cambridge University Press.
- Singer 2002: Peter Singer, *Unsanctifying Human Life: essays on ethics*, ed. Helga Kuhse. Oxford: Blackwell.
- Sober 1994: Elliott Sober, *From a Biological Point of View*. Cambridge: Cambridge University Press.
- Strawson 1962: Peter Strawson, 'Freedom and Resentment', in his *Freedom and Resentment and other Essays*, London: Methuen, 1974.
- Teichman 1996: Jenny Teichman, *Social Ethics*. Oxford: Blackwell's.
- Tooley 1972: Michael Tooley, "Abortion and Infanticide", in Singer, ed., *Applied Ethics* (Oxford: OUP, 1987).
- Vlastos 1981: Gregory Vlastos, "The Individual as Object of Love in Plato", in his *Platonic Studies*, Princeton, NJ: Princeton University Press, 3–42, 2nd edn.
- Warren 1997: Mary Anne Warren, "On the Moral and Legal Status of Abortion", in H.Lafollette, ed., *Ethics in Practice* (Oxford: Blackwell, 1997).
- Williams 1985: Bernard Williams, *Ethics and the Limits of Philosophy*. London: Penguin.
- Williams 2006: Bernard Williams, *Philosophy as a humanistic discipline*. Cambridge: Cambridge University Press.
- Williams 2006a: Bernard Williams, "Hylomorphism", 218-230 in his *The Sense of The Past*. Cambridge: CUP.
- Ludwig Wittgenstein, *Philosophical Investigations*. Oxford: Blackwell, 1951.

## Notes

---

<sup>i</sup> Anscombe mistranslates *zur Seele* as “to a soul”; I have corrected this to “to *the* soul”.

<sup>ii</sup> Perhaps we could do ethics entirely or almost entirely without using the word or concept “person”: so numerous philosophers have at times suggested, including Jenny Teichman (1996), Bernard Williams (2006a), Christopher Coepe (2006), and myself (I have at least had this thought; I can’t now remember if it has ever seen the light of published day; something close to that thought is certainly at work in Chappell 1998, Chapter 4). However, I don’t pursue that idea here.

<sup>iii</sup> Criterialism, as I call it, is pretty much the same view as what McMahan 2005, following Rachels 1990, calls “moral individualism”. This is the view (Rachels 1990: 173) that “how an individual may be treated is determined, not by considering his group memberships, but by considering his own particular characteristics”.

<sup>iv</sup> Or something very close to criterialism. I need not commit myself yet on whether these authors are defending necessary-and-sufficient-conditions versions of a criterial view.

<sup>v</sup> Something like the criterial view pops up in many other places. For instance, it seems to be behind Olson 2007’s suggestion that “person” is a phase sortal: “Being a person may be only a temporary property of you, like being a philosopher”.

<sup>vi</sup> It might be suggested that (some instance of) Criterialism is a *bespoke* view of personhood, aimed only at defining what persons are for some specific purpose. Such a position is perfectly possible, of course, but I think it is very clearly *not* what is at issue in discussions like the four I cite here, all of which are aimed at completely general conclusions about “what persons are”: what they are in their essence, and for *any* possible purpose. (Thanks to an anonymous referee for discussion of this point.)

<sup>vii</sup> For more on personhood and language, see Patterson and Gordon 1993.

<sup>viii</sup> David Hume: “For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch *myself* at any time without a perception, and never can observe any thing but the perception. When my perceptions are removed for any time, as by sound sleep, so long am I insensible of *myself*, and may truly be said not to exist” (*Treatise* 1.6.3). Derek Parfit: “We could give a complete description of reality without claiming that persons exist” (Parfit 1984: 212). The Buddha: “There exists no individual, it is only a conventional name given to a set of elements” (quoted in Parfit 1984: 502).

<sup>ix</sup> Autism involves very various sorts of emotional dysfunction. It can involve *over*-emotionality and “inappropriate” emotional responses as well as a lack of emotional response. But that just underlines the absurdity of using “emotionality” as a necessary condition of personhood. For then autistics who are over-emotional will count as persons, autistics who are “cold” and “withdrawn” won’t.

<sup>x</sup> There is a surprising contrast between some recent philosophers’ willingness to grant moral standing to “potential persons” in the sense of metaphysically possible future human beings who admittedly do not exist right now, and their reluctance to do so to “potential persons” in the sense of young human beings who admittedly do exist right now.

<sup>xi</sup> “The wrongness of killing us is understood in terms of what killing does to us. Killing us imposes on us the misfortune of premature death. That misfortune underlies the wrongness... The misfortune of premature death consists in the loss to us of the future goods of consciousness... What makes my future valuable to me are those aspects of my future that I will (or would) value when I will (or would) experience them... What makes killing us wrong, in general, is that it deprives us of a future of value. Thus, killing someone is wrong, in general, when it deprives her of a future like ours.” (Marquis 1997: 95-96)

<sup>xii</sup> And it is nearly always *the* criterion, i.e. some *one* criterion, that we are talking about. (Warren’s list of six criteria is the only exception to this monism that I can think of.) The suspicion that the value of persons might be multiform not monistic, that there might be not just one simple and readily-articulable reason for valuing persons but *all sorts* of reasons, some of them far from simple or easy to put into words, is not at all a characteristic suspicion for criterialists.

<sup>xiii</sup> About the argument that I develop from this point on, about “how we treat” people, it is natural to ask: are these descriptive or normative claims? The answer to that, as quite often with Wittgensteinians like (I suppose) me, is: a bit of both. What is on offer is a phenomenology of (part of) the moral life. My claim is that the most basic and fundamental truths about this phenomenology makes certain options in thinking about ethics that might seem attractive and viable, such as criterialism, actually incoherent, because of its mismatch with what is basic and fundamental. Because of how things *are*, there are some things we *shouldn’t* think or do: both a normative and a descriptive thesis.

---

Another worry that might occur to some here: am I attacking criterialism because it is a *theory* of personhood? If so, isn't it hypocritical of me to go on to develop what is really just another theory of personhood? The answers are No and No. I'm not attacking criterialism because it's a *theory*, but because it's a *rotten* theory. A good theory (or if you dislike that word, a good philosophical account) of personhood will make good sense of our intuitions; this criterialism signally fails to do.

(My thanks to anonymous referees for pressing these questions.)

<sup>xiv</sup> The charming comparison is Singer's: Singer 1993: 89-90.

<sup>xv</sup> But when exactly? As part of the child's fourth birthday celebrations? On going to primary school? Would a Singerian society have a ritual for an induction into personhood?

<sup>xvi</sup> At least, the term "objective" is unfortunate if it implies, as I suspect it does in passages like the one from Strawson that I cite in the next Note, that there is something *non-objective* about participant reactive attitudes.

<sup>xvii</sup> "We look with an objective eye on the compulsive behaviour of the neurotic or the tiresome behaviour of a very young child, thinking in terms of treatment or training. But we sometimes look with something like the same eye on the behaviour of the normal and the mature. We have this resource [the objective attitude] and can sometimes use it. Being human we cannot, in the normal case, do this for very long, or altogether. But what is above all interesting is the tension there is, in us, between the participant attitude and the objective attitude. One is tempted to say: between our humanity and our intelligence. But to say this would be to distort both notions." (Strawson 1962: 66-67, with omissions)

<sup>xviii</sup> "One speaks and expresses emotions to and not simply at one's cat. Nonetheless, although we of course discipline them, we don't press claims against or hold our pets accountable in the same way we do with one another... Granted, when I look into my cat's eyes, I can't get over the feeling that he is looking also into mine in some personal way or shake the hope that he is seeing me in some way other than 'the guy who feeds me'. But I find it utterly impossible to sustain the thought that he can imaginatively enter into my point of view or acknowledge me as a being with an independent perspective." (Darwall 2006: 43)

Pets are an interesting case of something like a "pretend" interpersonal relationship, if "pretend" is the right thing to call it. We know that, unlike infants, our cats will *never* become responsive in the way that persons do. That doesn't stop us (or at least the dedicated pet-owners among us) from extending something interestingly like the constitutive prolepsis even to cats, hamsters, or stick-insects. Children, of course, do something similar with their dolls or other toys—and here "pretend" clearly is the right word. For a recent fashion to which similar comments apply see <http://www.nytimes.com/2009/07/26/magazine/26FOB-2DLove-t.html> . (Thanks to an anonymous referee for drawing this to my attention.)

<sup>xix</sup> "Intrinsic" in the sense of "essential", I mean; not in the sense of "non-relational", obviously.

<sup>xx</sup> Gaita 1991: 117: "An animal can suffer, but it cannot curse the day it was born. An animal can be afraid, but it cannot be ashamed of its fear and despite itself. An animal can be happy, but it cannot be joyous. An animal cannot give of its substance to certain pursuits and be admonished for doing so. One could go on almost indefinitely. The problems of life's meaning cannot arise for an animal. Only a being for whom life can be problematic can have a spiritual life, and therefore have a soul."

<sup>xxi</sup> For more examples applying this argument see Hills 2010: 235-7.

<sup>xxii</sup> So I say. In the case of trees, of course, there was at least one stern Victorian moralist who thought otherwise, namely W.E.Gladstone, whose hobby was chopping down trees.

<sup>xxiii</sup> For this speculation deployed for a different purpose, cp. Kahane 2010.

<sup>xxiv</sup> Cp. Gerald the gorilla in the well-known *Not the Nine O'Clock News* sketch:

<http://www.youtube.com/watch?v=MpbMm0433I>

<sup>xxv</sup> "Suppose that the experiments attempting to teach language to chimpanzees had been successful beyond their originators' wildest dreams. The apes in question gained a complete mastery of English within a few years, were able to attend school and later university, and made many close friendships with human beings... in these imaginary circumstances it would plainly be absurd to claim that the apes lacked moral standing, or had a moral importance that was lower than our own." (Carruthers 1992, Ch.3)

<sup>xxvi</sup> Carruthers' own view about animal rights is much closer to speciesism than to animal liberationism. He holds that there is an indirect argument, arising from contractualism, for granting all humans and no animals moral standing (for a brief exposition see his online paper "The animals issue" at <http://www.philosophy.umd.edu/Faculty/pcarruthers/The%20Animals%20Issue.pdf>). Since I think it is obviously true that (non-human) animals do have moral standing in their own right, based roughly on the wrongness of causing unnecessary suffering, damage, and destruction, it seems I am closer to animal liberationism than Carruthers is.

---

<sup>xxvii</sup> Jeff McMahan (2005: 366) writes that his “moral individualism” gives us reason to deny that it is a grave misfortune for a human individual who is disabled to have limited cognitive capacities. After all, it is not a misfortune for an individual animal to have capacities of roughly the same level; and according to moral individualism, it is only the individual creatures that we should be considering. So either both the human and the animal are unfortunate, or neither is.

This seems a very counter-intuitive consequence of “moral individualism”. Surely it matters how the creatures got into their states: at the very least, there must be a difference in fortune between a human individual who was born at a low cognitive level, and one who is at that level because of some horrendous accident? So far as I can see, McMahan must deny this. For presumably he thinks that my past stands to me as a relational characteristic, not an intrinsic one—and the whole point of “moral individualism” is that only “intrinsic characteristics” count morally.

<sup>xxviii</sup> Bernard Williams’ delightfully mordant phrase: Williams 1985: 46.

<sup>xxix</sup> The best presentation I know of the evidence for Aristotle’s blindness on the subject of pity is in Gaita 1991.

<sup>xxx</sup> For more on this point see Chappell 2003, Grau 2010, Vlastos 1981.

<sup>xxxi</sup> “It is curious, but till that moment I had never realized what it means to destroy a healthy, conscious man. When I saw the prisoner step aside to avoid the puddle, I saw the mystery, the unspeakable wrongness, of cutting a life short when it is in full tide. This man was not dying, he was alive just as we were alive. All the organs of his body were working—bowels digesting food, skin renewing itself, nails growing, tissues forming—all toiling away in solemn foolery. His nails would still be growing when he stood on the drop, when he was falling through the air with a tenth of a second to live. His eyes saw the yellow gravel and the grey walls, and his brain still remembered, foresaw, reasoned—reasoned even about puddles. He and we were a party of men walking together, seeing, hearing, feeling, understanding the same world; and in two minutes, with a sudden snap, one of us would be gone—one mind less, one world less.” (George Orwell, “A Hanging”, *Adelphi Magazine* August 1931)

<sup>xxxii</sup> My thanks for their comments and encouragement to a large number of people, including Julia Annas, Alex Barber, Chris Belshaw, David Benatar, Peter Cave, Stephen Clark, John Cottingham, Christopher Coope, Christopher Grau, John Haldane, Rosalind Hursthouse, Tim Mulgan, David Oderberg, Tim O’Hagan, Charles Taliaferro, Chris Tollefsen, and Garrath Williams; an audience in Leeds at the Society for Applied Philosophy in 2009; and three anonymous referees for this journal.



## **On the very idea of criteria for personhood**

### *Abstract*

I examine the familiar criterial view of personhood, according to which the possession of personal properties such as self-consciousness, emotionality, sentience, and so forth is necessary and sufficient for the status of a person. I argue that this view confuses criteria for personhood with parts of an ideal of personhood. In normal cases, we have already identified a creature as a person before we start looking for it to manifest the personal properties, indeed this pre-identification is part of what makes it possible for us to see and interpret the creature as a person in the first place. And that pre-identification typically runs on biological lines. Except in some interesting special or science-fiction cases, some of which I discuss, it is human animals that we identify as persons.

## On the very idea of criteria for personhood

Meine Einstellung zu ihm ist eine Einstellung zur Seele. Ich habe nicht die *Meinung*, dass er eine Seele hat.

My attitude towards him is an attitude towards the<sup>i</sup> soul. I am not of the *opinion* that he has a soul. (Ludwig Wittgenstein, *Philosophical Investigations*, II, p.152, tr. G.E.M.Anscombe)

Love is conditioned by its object, but love also yields its object. (Gaita 1991: 122)

### 1. Persons: humanism, speciesism, criterialism

Any ethical outlook much like ours will take as central some *primary moral constituency* (PMC): some class of creatures who all alike, and all equally, share in the highest level of moral rights and privileges. Most philosophical ethicists use “person” to mean at least “member of the primary moral constituency” (whatever else they may also mean by “person”). In this paper, that is how I shall use the word too.<sup>ii</sup>

So what makes any creature a person, a member of the primary moral constituency? One view is what I shall call humanism:

**Humanism:** Being human is sufficient for membership of the PMC.

Humanism faces two very common objections. The first is that humanism is arbitrary discrimination, like racism, sexism, ageism; humanism, in Richard Ryder’s famous phrase, is *speciesism*. The second objection, which is really just the same idea taken a little deeper, says that humanism bases its demarcation of persons on a biological property which is morally insignificant. Membership of this or that species—just like membership of this or that race, gender, or age-group—is, we are told, a morally insignificant detail. It is not the kind of thing that we should expect any important moral distinction to be based on.

In theory these objections could be put by anyone who rejected humanism. In practice they are usually put by proponents of another theory of personhood, which I shall call criterialism:

**Criterialism:** Actual possession of the criterial properties is necessary and sufficient for membership of the PMC.<sup>iii</sup>

Here are four very well-known enunciations of criterialism<sup>iv</sup>:

I propose to use “person”, in the sense of a rational and self-conscious being, to capture those elements of the popular sense of “human being” that are not covered by “member of the species *Homo sapiens*”. (Singer 1993: 87)

Persons are beings capable of valuing their own lives. (Harris 1985: 16-17)

An organism possesses a serious right to life only if it possesses the concept of a self as a continuing subject of experiences and other mental states, and believes that it is itself such an entity. (Tooley 1972: 82)

[The six key markers of personhood are] (1) sentience... (2) emotionality... (3) reason... (4) the capacity to communicate... (5) self-awareness... (6) moral agency. (Warren 1997: 83-84)<sup>v</sup>

As here represented, the criterial view is an all-purpose<sup>vi</sup> view about what personhood is, and why it matters, which is undeniably simple (as I shall argue, much too simple). Despite this simplicity, it is obvious, when criterialism is stated as baldly as it is in these four statements, that it has radical implications. Most noticeably, it implies a sharp contraction of the circle of intuitive moral concern. If rationality is a necessary condition of counting as a person, a member of the primary moral constituency, then all children under a certain age will fail to count as persons. (This bullet is explicitly bitten by Tristan Engelhardt: “If being a person is to be [*sic*] a responsible agent, a bearer of rights and duties, children are not persons in a strict sense” (Engelhardt 1989: 120).) If self-consciousness is a necessary condition, then many mentally handicapped humans will fail to count as persons. If the capacity to communicate is a necessary condition, then Jean-Dominique Bauby, of *The Diving Bell and the Butterfly* fame, would have failed to count as a person after his catastrophic stroke—had he not worked out how to signal with his eyelid.<sup>vii</sup> (Can whether Bauby is still a person after his stroke really depend on contingencies about whether he can still communicate?) If persons necessarily have “the concept of a self as a continuing subject”, and believe that they are such selves, then David Hume, Derek Parfit, and most Buddhists are not persons.<sup>viii</sup> If you cannot be a person unless you are capable of valuing your own life, then you cannot be both a person and a committed nihilist about value (i.e. someone who finds himself compelled to believe that no such thing as value exists). If persons have to possess “emotionality”, then Mr Spock in *Star Trek* is not a person, and neither, perhaps, are some extreme autistics.<sup>ix</sup> And so on. Criterialists can make personhood sound rather exclusive; it can seem as hard to qualify for personhood as it is to make membership of the Country Club.

Of course, these problems are consequences of taking these “markers of personhood” as indicators of necessary conditions for personhood. A less extreme version of the criterial view might take each of the markers on some list, or some or other combination of the markers, as a sufficient condition of personhood: any creature that displays just one of these properties, or some large-enough variety of them, will count as a person. We might then want to argue about what counts as displaying a property, or possessing a capacity. Must I actually evince emotions to pass Warren’s emotionality test? How often, and how convincingly? Do I possess the capacity to communicate only when I have learned a language? Or just when I have learned to sign, or to get others to read my thoughts and feelings? Or do I have the capacity to communicate all along, just in virtue of being a member of a species that communicates, linguistically and in other ways?

Here we touch on familiar and long-running debates about potentiality versus actuality in our assessments of personhood. It is commonplace for criterialists of Singer’s sort to point out that potential prime ministers obviously do not have the rights of actual prime ministers, so that “potential persons”<sup>x</sup>, or those with the

potential for what Don Marquis famously called “a future like ours”<sup>xi</sup>, can hardly be treated as the same moral category as actual persons, or “us”. It is equally commonplace for the opponents of Singer-style criterialism to object that there must be something wrong with an emphasis on *actual* properties that makes it a live question whether one is a person while asleep or under general anaesthetic.

Clearly there are problems about at least some of the suggestions listed above about what the criteria of personhood might actually be. But not all those suggestions are implausible, and it would be easy enough, in principle, to weed out or refine the implausible ones. True, it is striking how little criterialists feel they need to do this; in practice, criterialists often seem perfectly happy to stick with what, intuitively, seem implausibly over-demanding criteria for personhood.

However, the plausibility or otherwise of the various suggested criteria is not my main concern here. Obviously enough, other criteria could be devised which did not have these consequences. The key difficulties for criterialism that I want to emphasise in this paper are three.

The first, which I can state very quickly, is about the argumentative method underlying criterialism. To put it at its simplest, it *looks* as if criterialists typically adopt their preferred criterion of moral significance because they think it explains moral intuitions which we have anyway. The criterion is adopted, in other words, because it doesn't have counter-intuitive consequences.<sup>xii</sup> But then, the same criterion is deployed, by the very same authors—Peter Singer and John Harris, for example—precisely to *produce* counter-intuitive consequences, such as the moral acceptability of killing small babies or the mentally disabled. There is, I think, a serious question as to what can possibly be going on here. It is not entirely obvious why building a moral theory out of one intuition or set of intuitions that we certainly have, and then using this theory to attack other intuitions that we also certainly have, is anything more than arbitrary selectiveness. Perhaps this is a question that all system-building moral theory faces.

The second difficulty for criterialism, which I shall also mention only very briefly, is that criterialism can seem to imply a quite implausible picture of moral reasoning. Are we really to suppose that what we do when we come across any being we haven't met before is *first* see whether it satisfies our criterion for personhood, and *then*, if it does, start treating it as a person? One is reminded of that well-known caricature the Cartesian Detective: we start with the hypothesis—the *opinion*, as Wittgenstein says in my epigraph—that the new being is a person, and feel justified in acting in line with that hypothesis, just insofar as the behavioural evidence confirms it. If criterialism is to be at all plausible, it cannot in this way be a general view about our typical real-time interaction with others. Such a picture might apply to some very exceptional cases, some of which we will discuss in section 4b. But in general, criterialism will have to be a view about the *criterion of rightness* for taking other creatures to be persons, not about *our decision procedure* for assuming that they are persons. For quite obviously, our normal interaction with others typically involves no such Cartesian-Detective rigmarole, and it takes a good deal of philosophical callousing of the soul not to find something absurd, and morally repugnant too, about the idea that it even could. But if in general criterialism only states the criterion of rightness, and not our decision procedure, for taking other creatures to be persons,

then it still faces a further objection, my third—which it is the main purpose of this paper to develop.

This third objection to criterialism, on which I shall spend the rest of this paper, is connected to this second point. My claim is that the sort of properties that criterialists home in on are not criteria of personhood at all. Rather, they are dimensions of interpretation of beings that we *already* take to be persons.

For my own part, I have always been aware of at least a vague sense of discomfort and unease here, the moment we start talking about these (or any other) properties as if they were criteria for personhood of the sort the criterialists have in mind. There is a hard-to-articulate feeling that, despite the considerable first-blush plausibility of criterialism, it leaves out something crucial; a feeling of the kind that is naturally expressed in ordinary conversation by a trailing-off “Yes, *but...*”. I suspect that my unease is not unrelated to the unease that Raimond Gaita is bringing out when he writes (Gaita 1991: 115) that it is

naïve to think that the kinds of properties which interest philosophers when they ask what is a person, and which we often share to some degree with higher animals, play the kind of role in our treatment of one another which is assumed by those who argue that differences in treatment must be justified by relevantly different properties of those kinds. That philosophical perspective from which we are encouraged to reassess our sense of how we might justifiably treat animals distorts and indeed cheapens our understanding of human life.

Gaita 1991 is a marvellous exploration of one philosopher’s attempt to make articulate that trailing-off “Yes, *but...*”. This essay offers another philosopher’s. My attempt is not necessarily the same as Gaita’s, but not necessarily in competition with it either.

And it is perhaps worth saying before I begin that what I am trying to do here, what Gaita was trying to do, is *hard*: hard to articulate, hard to state clearly, hard to see in the round, hard to apply. Again and again criterialism wins out in philosophers’ discussions of personhood because, conversely, it is dead easy; it is such a simple and straightforward view of personhood. There are plenty of spurious analogies between science and ethics around, but at any rate we can trust this one: no more in the case of the nature of the person than in the case of particle physics does the fact that a view is *simple* in any way improve its chances of being *true*.

## **2. The proleptic view of personhood**

Contrary to what criterialism seems to suggest, we do not look for sentience or rationality or self-awareness in a creature as a test to decide whether or not that creature counts as a person. It’s the other way round. Having once decided, on other grounds, that a creature *is* a person, we know that this makes it the kind of creature which is likely to display sentience, rationality, self-awareness and the rest of the personal properties. Hence we look for displays of these properties from the creature. That is to say, we treat it as a person *in advance* of any such displays.<sup>xiii</sup>

The biggest and clearest example of this is an activity that is absolutely central to human life: parenting. Parents are, of course, aware of the differences in rationality, linguistic capacity, self-representing ability, and so forth between young human beings at various ages and developmental stages. Nonetheless, a parent's attitude towards her child is always, basically, what Wittgenstein famously calls "an attitude towards a soul". She does not do what criterialism (especially if understood in decision-procedure terms) might seem to imply she should do: start out by treating her child as an inanimate object, like a sofa or a fridge or a rubber-plant, and grudgingly consent to adjust her attitude to it, one little step at a time, only as and when it proves itself more than inanimate by passing a succession of behavioural tests for at least having interests like a snail's perhaps,<sup>xiv</sup> and then later on for the interests of some more advanced kind of animal such as a puppy, and eventually for full criterialist personhood.<sup>xv</sup> If parents *did* treat their children in this almost behaviouristic fashion, the parents would be callous monsters, and the children would be basket-cases. In a world where parents generally accepted this sort of criterial view, and applied it directly in their parenting practice, even the best-adjusted of us would be a wolf-child.

Fortunately, however, parents are not criterialists, either in their decision procedure or in the criterion of rightness for ascriptions of personhood that they actually employ. Rather, a parent treats her child from the very beginning—and from before it is literally and actually true—as a creature that can reason, respond, reflect, feel, laugh, think about itself as a person, think about others as persons too, and do everything else that persons characteristically do. From the beginning her attitudes towards the child are not only "objective" (to use Strawson 1962: 67's slightly unfortunate term<sup>xvi</sup>): they also include what he famously calls "participant reactive attitudes", of just the same sort as she adopts towards anyone else.

Of course, as Strawson implies<sup>xvii</sup>, these participant reactive attitudes' scope may be more restricted in the case of a very small baby. The point is that they are there; and that like the baby—often indeed in advance of the baby—they *grow*. As Alasdair MacIntyre notes in a recent discussion of the development of practical reasoning in the individual person (1999: 90), ethically and psychologically adequate parenthood centrally involves "a systematic *refusal* to treat the child in a way that is proportionate to its qualities and aptitudes" (that is, its actual and non-idealised qualities and aptitudes, the kind of qualities and aptitudes that interest the criterialist).

The parent who says to her three day old baby, "What do you think? Should we give you some more milk now?", or to her three year old toddler "How kind of you to share your toys with your sister when she's sad!", is not deluded about a neonate's capacity for interpersonal practical reasoning, or a toddler's capacity for calculatedly empathetic altruism. But she is not making a sentimental joke either. She is treating the baby or the toddler proleptically, in the light of the ideal of personhood. By years of treating her children as creatures who "have the personal properties"—in the sense that interests the criterialist—she makes it true that they *are* creatures who have the personal properties in just that sense.

The child staggers across and plonks a book down on his sibling's lap; the parent's reaction is "How kind of you to let her share your book!" Is that, in fact, what the toddler was doing? The right answer to this question can be: "Yes it is—once the

parent has given this reaction.” The parent’s reaction is an interpretation of the child’s deed. Before the reaction, perhaps, there was no fact of the matter about what the little boy was doing: the child did not know himself what his action was. For all he knew about it, his action might as well have been simply a random sequence of bodily movements. But now that his mother offers her interpretation, the child, on his mother’s authority, learns to see himself a certain way. Because his mother frames his act as one with a certain meaning (as the giving of a gift), and because the child sees his mother as seeing his act this way, and because his mother sees the child as seeing her as seeing his act this way... (Grice 1957), the act *comes to have* that meaning.

Paul Grice’s famous work on meaning and intention is even more relevant to another clear example of this sort of prolepsis, namely language-learning. A baby begins by babbling—that is, by producing all the sounds the human mouthparts can make. Some of these sounds get a response because they are sounds that occur in the parents’ language, while others don’t because they are not. (The inability of many English adults to pronounce the Scots *ch* in *loch* or the Welsh *ll* in *Llanfair* is a learnt inability, not mirrored in English babies, who can and do make both sounds until conditioned not to by English adults.) The baby “homes in” on the sounds, and then on the patterns or combinations of sounds, that get a response. Then it learns to correlate particular patterns of sounds with particular contexts by the same homing-in process. What the parents say is “She’s learned to say ‘tiger’” (my own first child’s first word). What has happened is that the parents, by treating their daughter *as if she has* the word for, and then the concept of, a tiger, have made it true that she *has* that word. (And, a little later perhaps, that concept too; though I am not suggesting that any one word or concept can be learned on its own: in Wittgenstein’s familiar image, “light dawns gradually over the whole” (*On Certainty* 141).)

Stephen Darwall too talks of prolepsis when describing a third connected process, namely the process whereby we come to be, and to hold others, morally responsible (Darwall 2006: 87-88):

There seem to be many cases where we wish to hold others accountable though we seem to have very good evidence that they are not free to act on moral reasons in the way our practices of holding someone fully responsible seem to presuppose... In some instances, for example with children, we seem simultaneously to move on two tracks in the process of inducting them into full second-person responsibility, sometimes treating them proleptically as though they were apt for second-personal address as a way of developing moral competence while nonetheless realising... that [at their developmental stage] this is an illusion.

Or an illusion, at any rate, if you agree with the criterialist about what it takes to be “apt for second-personal address”, and/or about what it takes to “have the personal properties”. My point, with which Darwall might or might not agree—I’m not sure—is that what it takes does not have to be already-realised and –actualised capacities of the sorts that interest criterialists. (The mother who asks her neonate child “Shall I give you some milk?” is certainly engaging in second-person address, *despite* the manifold incapacities, perfectly well-known to her, of neonates.)

There is at least this much truth in Dennett 1987's well-known idea of the "intentional stance": persons constitute each other as persons, both as agents and as patients, by *treating* each other as persons, as agents and patients. Something like the Davidsonian "principle of charity" (Davidson 1980) is at work in our mutual interpretations. By charitably, and proleptically, interpreting the other as a person, I *make* him a person.

The looseness and ambiguity of the term "person" is evident when we say things like this. In ordinary language, "person" can certainly, at one end of its semantic range, mean something like what Singer and other criterialists want it to mean, a "rational and self-conscious being" (or whatever). Yet at the other end of its semantic range, the ordinary-language term can also be a synonym for "human animal". It seems to be part of the *point* of our person-concept to allow those young humans who are persons only in the "human animal" sense to enter into a continuum of meaning towards the far end of which they can become persons in much stronger senses.

Many things in our shared life can be seen in this sort of way as aspirations towards an ideal which, like any ideal, is never fully realised. The basic sort of idealisation, the sort that grounds attributions of personhood, is not the only sort. We ascribe the rights to freedom of assembly and freedom of speech to every human being, even though we know that small babies can't speak and don't assemble (not without help), and that acutely asocial loners may well never (or at least never again) actually exercise either right. The behaviour of parents, besides (as described) involving the person-constituting prolepsis in a particularly clear form, is also full of other similarly-patterned proleptic idealisations that might, on a criterialist's view, be regarded as quite unrealistic. Parents give their children the opportunities to play the cello, to learn Spanish, to use their monthly allowances sensibly, to be polite to Great Aunt Maud (...), even though the parents may be fully aware that every one of these opportunities will in all probability be spurned. In minimally decent jurisdictions the treatment even of hardened criminals displays the same pattern of basic *openness* to the convict, however much of an unrealistic idealisation this openness may seem to involve: the jurisdiction goes on offering the criminals opportunities to reform, even though it is as good as certain that these opportunities will not be taken up. (I think that something like this refusal to "close the door" finally and irrevocably on the criminal is also, in connection perhaps with some thoughts about the state's obligation to be fundamentally benevolent to the citizens for whose sake alone it after all exists, the ground of a good reason to reject capital punishment. But that's another story.)

In general, the idealisation that accords X the status due to a  $\phi$ er does not have to involve any kind of expectation that X will ever actually  $\phi$ , or even can  $\phi$ . What it *does* have to involve is the understanding that X is of a kind that characteristically includes  $\phi$ ers. The human kind includes individuals who characteristically (at the right point in the human life-cycle) will freely assemble and speak. That is why it is not ridiculous to give the rights to free assembly, speech, and political action, e.g. becoming a senator, to any and every human, including asocial misanthropes, those (as yet, or as of recently) physically incapable of speech, and those who choose never to utter a word. Contrast the equine kind, which does not characteristically include any such individuals. That is why it *is* ridiculous to give these rights (or the status of a senator) to any horse.



Considered as a kind, children characteristically include individuals who learn the cello or Spanish or wise use of money or good manners to venerable relatives (well... eventually). That is why the parents who try to teach these lessons to their children are not acting absurdly even if they know in advance that *this particular* child is not going to respond—whereas parents who tried to teach these lessons to the family goldfish or the rubber plant *would* be acting absurdly. Similarly, there is nothing absurd about offering the old lag a place on the rehabilitation programme, even if it seems certain that *this* old lag will always refuse or abuse the offer: and the reason why not has to do with the general fact that some old lags do sometimes reform.

*A fortiori*, there is nothing absurd about giving the rights of free speech, assembly, private choice and the rest of it to women, or to other races, or to homosexuals or transgender people, or to the disabled, or to gingers, or to the various other sorts of historically marginalised groups. People in these categories are not another kind, as horses and hamsters are. Their well-being is *human* well-being, just as white male heterosexuals' is, and just as horses' and hamsters' well-being is not. That, in a nutshell, is why speciesism is *not* like racism or sexism.

Clearly, then, this is not just a point about the induction of babies and very small children into the social life of humanity at large. All humans, even the healthy intelligent mature well-adjusted independent affluent adults that typical criterialism is so clearly focused on—the Country Club again—are always only incompletely and impurely agents. Incompletely, because we are never all agents can be. However well things go, we never do or even can reach the full potential of human well-being, which would include a fully worked-out articulation of our own agency, a full coordination and marshalling of our own forces and powers for action in the world: towards this we are always only, at best, on the way. And impurely, because each of us always *can* be interpreted, by a sufficiently determined sceptic, as not really an agent at all, but a victim of the instinctive or other determining external forces that constantly threaten anyone's agency. To extend “an attitude towards a soul” towards *anyone* is always, and not just with children, an idealisation. As Tamar Schapiro notes within a Kantian framework (Schapiro 1999: 723):

[T]here is a sense in which no one, regardless of age or maturity, is able to achieve autonomy on Kant's view. This is because the notion of autonomy in Kant is an ideal concept which outstrips all possible realisations in experience. Strictly speaking, every instance of human willing is necessarily an imperfect realisation of transcendental freedom, and every virtuous character necessarily falls short of perfect virtue. And yet the applicability of the moral law depends on our mapping these ideal concepts onto ourselves and one another for the purposes of guiding action. So we are to regard the social world as a community of autonomous agents despite the fact that perfect realisations of autonomy are nowhere to be found.

The criterialists are right to insist that properties like self-awareness, emotionality, rationality, and so forth are crucial parts of our concept of a person. It does not follow that these properties can be used as the criterialists want to use them: as *tests* for personhood. Even at the cool-hour level of the criterion of rightness—never mind the

real-time level of deliberative procedure—to treat someone as a person is not to put a tick in the box by her name, to show that she has passed some inspection or met some standard, of rationality or self-awareness or emotionality or whatever. Indeed, it seems no less arbitrary discrimination to say “Sorry, you’re not rational or self-aware or linguistically capable or emotionally responsive enough to count as a person” than to say “Sorry, you’re not white enough to count as a person”. Behavioural properties like rationality, self-awareness, emotionality are not tests for, but parts of the ideal of, personhood. To treat someone as a person is to engage with him as the kind of creature to which that ideal applies. So to treat him is not, at the deepest level, a response to his *behaviour* at all, but to his nature. To see some creature as a person is to take an attitude to that creature which, *before* any behavioural evidence comes in, is already different from our attitudes to creatures that (we think) aren’t persons. This is the second-person standpoint (to use the title of Darwall’s outstanding book): the attitude we take to persons, and do not take to sofas, fridges, or rubber plants.

The fact that it takes charity, in more than one sense, to constitute persons imposes on persons a crucial sort of vulnerability and dependence: “The personal reactive attitudes rest on, and reflect, an expectation of, and demand for, the manifestation of a certain degree of good will or regard on the part of other human beings towards ourselves” (Strawson 1962: 70). All human persons are, in Alasdair MacIntyre’s splendidly apt phrase, “dependent rational animals”, as MacIntyre himself emphasises (MacIntyre 1999: 81-85):

In most moral philosophy the starting point is one that already presupposes the existence of mature independent practical reasoners whose social relationships are the relationships of the adult world. Childhood, if noticed at all, is a topic that receives only brief and incidental attention... To become an effective independent practical reasoner is an achievement, but it is always one to which others have made an essential contribution. The earliest of these relate directly to our animal existence, to what we share in our development with members of other intelligent species... [But what] we [also] need from others, if we are not only to exercise our initial animal capacities, but also to develop the capacities of independent practical reasoners, are those relationships necessary for fostering the ability to evaluate, modify, or reject our own practical judgements... Acknowledgement of dependence is the key to independence.

We are vulnerable to each other, and dependent on each other, because it is remarkably easy to attack persons by withdrawing the second-personal attitude from them. Part of the reason why it is so easy to stop seeing others as persons is, as I said before, because the interpersonal attitude always involves a degree of idealisation. There is always *some* truth in sceptical or cynical views that reject that idealisation. (For example, the human person *is* a physical object—even if it is not “merely” a physical object.)

We fail to see others as persons in a small way when we explain away what others do, or refuse to listen to them or to take their wishes and ambitions seriously. (Parents are never perfect in charity; children’s protests against such treatment are a recurring theme of family life.) We do it in bigger ways when we “use” others, or (in Kant’s phrase) fail to treat them as “ends in themselves”. At the limit, we can withdraw the interpersonal attitude from others completely, by pretending that they

simply aren't there, as British and Irish settlers did to the aboriginal Australians under their notorious legal doctrine of *terra nullius* (see Gaita 1998). Or we can do it by treating them merely as physical objects, as often happens in pornography, violence, and murder. Or, lastly, we can do it by declaring them "non-persons", as the Nazis did to the Jews—and as criterialists routinely do to the very old, the very young, and the very disabled.

The personal properties, then, set the content of the ideal of personhood. They do not set the boundaries of the class of persons (the PMC, as I've also called it). What does set the boundaries? That is my question in section 3. My answer will be the humanist answer already advertised in section 1: roughly, the boundaries of the class of persons are the boundaries of the human species.

### 3. The truth in speciesism

Rubber plants, fridges, sofas, human babies: we come to the last of these four sorts of things with a very different attitude from our attitudes to the other three sorts. And my point is that we *come to them* with different attitudes. Our attitudes to them are not responses to evidence that we gather *after* we have "come to them". *The moment I meet* a rubber plant or a fridge, I am perfectly sure, unless my sanity is in question, that it will not attempt to talk to me or tickle me—as, say, a baby might.

What makes me so sure of that? Well, simply the sort of thing a rubber plant or a fridge is, and the kinds of behaviour that it is natural to expect from that sort of thing. Rubber plants and fridges are not the kind of thing that tries to talk to you or tickle you. Babies are. And if you ask "What kind of thing is that?", the obvious answer is: "Young human beings". If some creature is a human being, it is completely natural to expect it to be the kind of thing that might (at least if it's a family member) try to talk to you, or tickle you, or otherwise demonstrate its (budding) mastery of the kinds of mental capacity that criterialists take to be criteria. (Talking to someone and tickling someone both require interesting interpersonal capabilities. To do either, you have to want things to be a certain way for someone else—roughly, you have to want to get them to think a thought, or to feel a tickle. These are wants with quite sophisticated contents.)

This sort of expectation is based squarely on the nature of the creature in question. And that means, on its nature *as a member of the human species*. Our understanding of which creatures are persons, and so may be expected to exhibit the personal properties, is not formed on an individual basis but on the basis of generalisation from experience of humans in general. I have already mentioned Wittgenstein's remark that "my attitude towards him is an attitude towards the soul"; on the same page of *Philosophical Investigations*, Wittgenstein also says that "the human body is the best picture of the human soul" (*der menschliche Körper ist das beste Bild der menschlichen Seele*). And perhaps what he means is this: that when we see that some creature is a human, just seeing that is enough to license us to adopt the interpersonal attitude to that creature.

We know in advance how we can expect any particular rubber plant or fridge to behave, because we can draw on a bank of past experience (our own and others') of

rubber plants and fridges. In this sense, the fact that a thing is a rubber plant or a fridge sets the scope of our expectations about it. In just the same way, we know what kinds of capacities and properties humans in general have. It is that background knowledge that we draw on, whenever we apply the interpersonal attitude to any individual human being, simply on the basis that it *is* a human being.

We obviously do the same with other species of animals. When I meet a cat or a gerbil, I treat it neither as an inanimate object like a sofa or a fridge or a rubber plant, nor as a person. (Not, of course, that I treat all inanimate objects the same, either. Sofas, fridges, and rubber plants are very different things, especially rubber plants.) You can play games with a gerbil. With a dog or a cat, you can even have a cuddle (in the reciprocal sense, I mean).<sup>xviii</sup> Neither is possible with a fridge. On the other hand, you would be wrong to expect a dog to play a good hand of contract bridge, or a cat to give you good investment advice, or a gerbil to read your mind, as a human being might. What you expect of an individual dog or individual cat will be different from what you expect of an individual gerbil, and both will be different from your expectations of inanimate objects. In each case, your expectations will be set by your background knowledge of the different kinds in question. Where dogs, gerbils, cats, and indeed all other kinds of living thing are concerned, this obviously means the different *species* in question. Why should it be otherwise with humans?

These remarks should enable us to see—*pace* the criterialists—how central it is, to our moral responses to the things and creatures around us, to assign them to species. It is species-assignments that enable us to answer what you might reasonably think is the most basic question of all about any thing that might be in front of us at any time: namely, what kind of thing it is. Until we know what species it is, we do not know how it may act, or how to treat it: what is good for it, how we might harm or hinder it, what counts as health for it, and so on. The nature of the creature determines what well-being is for it; and its nature is its species. To put it another way, the best way to answer “What is it?”, when this question is asked about any individual animal, is to give the name of its species. Or as we might ask moral individualists like McMahan: “If species-membership isn’t an intrinsic property, what *is*?” Singer, Rachels, and McMahan never tell us why *being human* cannot count as any creature’s “own particular characteristic”. It can’t be merely because *being human* is a group membership (as McMahan 2005 suggests). For there is nothing to stop a given property from being both a “group membership” and an “own particular characteristic”; that is how it is with *being a vertebrate*, for instance. To put it more formally, nothing stops a given property from being both a relational property, and also an intrinsic property.<sup>xix</sup> Indeed, if Kripke is right in his well-known thesis of the metaphysical necessity of origin, such properties must be very common.

This shows how humanism can be true without falling prey to a charge of arbitrary speciesism. To treat any species of creatures as persons is to treat them, not as successful passers of some test for the personal qualities, but on the assumption that they are individuals of a species the ideal for which is (roughly) the highest attainable development of those qualities. In this sense, references to species are an ineliminable part of our talk about persons.

In that sense, but not in some others. We are now in a position to respond to four familiar objections to “speciesism”. I consider these in section 4. The aim of my

response is not of course to defend speciesism, if that is defined as “*arbitrary* moral discrimination on the grounds of species”. But it is to defend humanism, the view that not all moral discrimination on the grounds of species *is* arbitrary.

#### 4. Four objections to speciesism

*a. All animals are equal.* Peter Singer (1993: 57) writes that “No matter what the nature of the being, the principle of equality requires that the suffering be counted equally with the like suffering of any other being”. Nothing I have said so far commits me to rejecting this application of the principle of equality. If I were to accept it, I would want to add, first, that plenty of other things besides suffering “count”, and secondly, that however alike human and animal *pain* (the physical sensation) may be, human and animal *suffering* (the state of mind) are mostly quite unlike. A cat can be tortured like Gloucester, but it cannot suffer like King Lear.<sup>xx</sup> And suffering matters more than pain.

But perhaps we should reject this application of equality anyway. Right now there is (quite likely) monstrous pain, and a fair amount of suffering too, on the plains of the Serengeti, where a wounded buffalo is being eaten alive by a pride of lions. Singer’s principle of equality requires us to say that our duty to prevent the buffalo’s pain and suffering in the teeth of the lions is directly proportional to the amount of pain and suffering involved. If any implication of any moral theory is absurd, this one is. There just isn’t any duty, not even a *prima facie* one, for us to act as Zoological Pain Police. (Don’t say: “We can’t intervene, because the lions might eat *us*”. Where is your moral courage? And don’t say: “But it causes less pain in nature overall if we don’t intervene in nature in this sort of way.” As many writers on the theological problem of evil are keen to remind us, most animal suffering is gratuitous. If lions can be caught and trained to jump through hoops, perhaps they can be caught and trained to kill their prey cleanly before eating it. So is anyone *trying* to train them? Of course they’re not.) The vast majority of the huge amount of animal pain and suffering that happens every day is simply none of our business. “It is not an accident or a limitation or a prejudice that we cannot care equally about all the suffering in the world: it is a condition of our existence and our sanity” (Williams 2006: 147).<sup>xxi</sup>

So it looks like Singer’s principle that all suffering counts for the same, no matter what the species involved, is just false. On the other hand, some weaker principle of this general sort seems clearly true: perhaps “Animal suffering is bad for the same sort of reasons as human suffering is bad”, or something like that. What also seems clearly true is that humans should do what they can not to generate animal suffering, and to prevent animals from being needlessly hurt or damaged *by humans* (and sometimes, though less often, by accident or by each other). This is a much weaker principle than Singer’s equality-of-suffering principle. But perhaps even this is strong enough to generate a good argument for vegetarianism. I suspect it all depends what counts as needless hurt or harm.

Hence “the limit of sentience” is not “the only defensible boundary of concern for the interests of others” (Singer 1993: 58). Certainly that limit marks one distinction between the ways we should take an interest in the welfare of other creatures. For example, hacking a frog to pieces is wrong in a way that hacking an

orange to pieces is not, partly because an orange certainly cannot feel pain, and a frog pretty definitely can. But only partly. Quite aside from the pain involved, destroying a living creature for fun is wrong in a way that destroying a piece of fruit for fun is not. (Hacking trees, or anaesthetised frogs, to pieces for fun is usually wrong too.<sup>xxii</sup>) Unless we think that nothing matters except pain and pleasure—and this is not, e.g., Singer’s position either (Singer 1993: xi)—there is no reason to think that the distinction between sentience and non-sentience is the only fundamental moral distinction we can make among beings; nor that it is the most important.

There are some interesting complications about the notion of sentience. If “sentience” means simply “capacity to feel (sensational) pain and pleasure”, then sentience does not fit the story that I have developed about how the personal properties come into being. The capacity to feel pain- and pleasure-sensations is certainly one of the capacities that we expect to find in any human person. But it is not, unlike rationality or intentionality, a capacity whose presence in any person is even partly constituted by others’ idealising assumption, expectation, or hope that it is there. Whether or not a being can feel pain does not depend on others’ adopting “participant attitudes” or taking the “intentional stance” towards him, in the way that (I have argued) his capacities for self-consciousness, rationality, and suffering do. I don’t think this shows that sentience is a counter-example to my thesis about how the person-constituting properties are developed. Rather, I think it shows that sentience is not one of the person-constituting properties. A creature which had the (other) person-constituting properties, but lacked the capacity for pain- and pleasure-sensations, would still undeniably be a person.<sup>xxiii</sup> A creature which had the capacity for pain and pleasure, but lacked the (other) person-constituting properties, surely would not be.

*b. Aliens, talking chimps, Locke’s parrot, and a possessed fridge.* The position that I am defending under the name of humanism says that, normally, to be a person it is *sufficient* to be a human being. Not *necessary*, because humanism does not imply that members of other species could not be drawn into the moral community of human persons. On the contrary, I can see at least three distinct ways in which this could happen. One is that we could discover a whole species—aliens, or angels, or what have you—which is like the human species in that its members are characteristically persons (in one sense or another). A second is that a species already known to us might, *en masse*, somehow change so as to become characteristically persons. Or, third, some particular individual creature—Locke’s parrot, say (Locke, *Essay* 2.27), or Washoe the chimp (Singer 1993: 111)<sup>xxiv</sup>, or some group of creatures, like Peter Carruthers’ Chimp Class of ’92<sup>xxv</sup>—might start to count as persons, perhaps e.g. by displaying the personal properties, even though this behaviour is otherwise completely untypical of the species in question. (Come to that, I suppose a possessed fridge might, in principle, start displaying the personal properties.)

In these special cases, we perhaps *will* have to appeal to something like the criterial view of personhood as our decision procedure for the vexed cases. Apparently we will have to ask whether this alien species, or this new class of highly-educated chimps, or this individual chimp, parrot, or spooky fridge, displays something like the criterialist’s personal properties, *before* we can decide whether to count these beings as persons.

Doesn't this show that my arguments against criterialism were mistaken? Not at all. It does show that something like criterialism may perhaps be our decision procedure if and when we encounter a being *that we do not already know how to classify as a person or not a person by way of its species*. (Perhaps; actually I doubt even this much is true, but at any rate it is the most that is true.) However, it would be utterly misleading to generalise from thought-experiments about these special and rare cases—almost all of which, to date, are imaginary—to alleged conclusions about the normal cases. (It would be as misleading to do that as it would be to generalise the justificatory/ rational structure of *every* belief we have about the external world from the justificatory/ rational structure of the external-world beliefs that we might form when we are thinking about the classic, and nearly always imaginary, sceptical predicaments.) Unfortunately, just this generalisation is a criterialist stock-in-trade. The most famous example of it is perhaps Mary Anne Warren's derivation of the personal properties (her list was quoted in section 1) from an imagined meeting with aliens. But even if Warren is right that this is the decision procedure that we would have to adopt with *aliens* to settle whether they were persons in some extraordinary imaginary scenario, nothing at all follows about our beliefs (either at the decision-procedure or at the criterion-of-rightness level) about whether any *humans* are persons in a multitude of ordinary real scenarios. For we have met humans before.

But anyway, I suspect Warren's description is much too black-and-white to get anywhere near characterising the decision procedure that would really be involved in such a meeting. (This is why I say, above, only that we would probably have to use *something like* the criterialist approach in alien encounters and similar.) For one thing, there is not the slightest reason to think that, in such a meeting, *yes* and *no* would be the only possible answers available to the question "Are these persons?". Even *Star Trek* yields a rather subtler answer—"It's life, Jim, but not as we know it"; the world, after all, is a complicated place, and it constantly throws us surprises. For another, our responses in such situations would be—or should I say *will* be?—based on all sorts of factors and reactions, most of them probably defying explicit articulation; on judgements about "forms of life", in fact. It would not and could not be based on some neatly finite check-list of diagnostic properties like Warren's. Looking beyond *Star Trek*, science fiction is full of wonderfully complex and subtle explorations of such encounters. The contrast with the mechanical simplicities of a criterialist list such as Warren suggests is instructive—and not *entirely* to be dismissed as unfair on the grounds that Warren is, after all, not writing science fiction.

*c. The two-species scenario.* A third interesting objection to speciesism is offered by Peter Carruthers:

Suppose it had been discovered that human beings in fact consist of two distinct species, otherwise hardly distinguishable from one another, the members of which cannot inter-breed. In these circumstances it would plainly be objectionable for the members of the majority species to attempt to withhold moral rights from the members of the minority, on the mere ground of difference of species. This, too, would be obvious speciesism. (Carruthers 1992 Ch 3)

Carruthers, I think, is quite right to claim that it would alter nothing morally speaking if we discovered that "the human species" was in fact two different

biological species—perhaps *homo neanderthalis* and *homo sapiens*—living together in a single moral community. (Modulo the point about not interbreeding, some anthropologists think that this has actually happened: Trinkaus and Shipman 1993.) Pace Carruthers, his argument does not show that there is no species-classification that bestows participation in the moral community on its members, and so that species-classifications are morally irrelevant. What his argument shows is that there may be *more than one* species-classification (or other kind-classification) that bestows participation in the primary moral constituency on its members. That is something that we should grant anyway given the possibility of other species whose members are characteristically persons; and it is not inconsistent with the claim that species-classifications are morally significant.<sup>xxvi</sup>

*d. Species and Darwinism.* A fourth objection to my argument is the claim that species don't exist: that as a matter of strict biology, there is no clear differentiation between humankind and "other species", only a differentiation between different parts of the same evolutionary family tree. Species therefore are populations, not Aristotelian (or Kripkean) natural kinds (Sober 1994). True, but irrelevant to my argument, which can work equally well with the population conception of species, or indeed with any conception at all that will allow "humanity" to count as the name of a kind of *some* kind. At least for creatures like us, whose temporal experience runs only over decades, hardly ever over centuries even, let alone the millions of years that speciation takes, it should not be a controversial thought that a population of creatures sharing a common genetic heritage, physiology, and ethology can sensibly be treated as a unitary grouping, as in fact common sense treats it. And that thought is all I am committed to meaning by "species".

So it looks like humanism can resist these four objections. And it looks like humanism can accept, indeed welcome, the thought that humans are not the only persons, the only members of the primary moral constituency—or at least, they might easily not be. What about the question on the other side: the question whether all human beings are persons? I turn to this question in the next and final section.

## **5. Are all human beings persons?**

Humanism as I have developed it here straightforwardly implies that all human beings are persons. Our treatment of any human being should be conditioned by the background of expectations, hopes, and aspirations that spell out what we know, from experience, humans in general can be. *Eudaimonia* in its broad outlines is the same for all human beings; and requires, as a general rule, that we must give all human beings the space to achieve *eudaimonia*—whether or not they predictably *will* achieve *eudaimonia*. To deny this space to any individual human being is to exclude that individual from the moral community of persons. And that is a serious injustice.<sup>xxvii</sup>

The very young are not excluded from the moral community of persons, just because they have not yet achieved many key forms of human *eudaimonia*; the mentally handicapped are not excluded from the moral community of persons, just because they cannot ever achieve many key forms of human *eudaimonia*. As I pointed out in section 1, *all* persons are incomplete and impure in their agency. No one ever



does everything she could with her own nature (partly, of course, because there are just too many things she could do). In one way or another, *all* of us are less than fully free to pursue happiness; everyone is wounded or damaged or limited or compulsive, or just plain ill, in some respect. That does not make it wrong to treat “normal” humans, whatever their age or IQ, with an interpretive charity, and with a kind of pity, that sees beyond their limitations, and leaves the door open, in each individual’s case, to what *any human* might become—even if we are well aware that *this person here* has no chance of becoming that.

[T]he kind of care that was needed to make us what we have in fact become, independent practical reasoners, had to be, if it was to be effective, unconditional care for the human being as such, whatever the outcome. And this is the kind of care that we in turn now owe or will owe. Of the brain-damaged, of those almost incapable of movement, of the autistic, of all such we have to say: This could have been us. (MacIntyre 1999: 100)

There is the possibility of deep moral concern for retarded people, in which they are seen as having, however incomprehensible we may find it, a human fate, as much as anyone else’s.... Someone may be very touched by the response of a severely retarded person to music; and there may be in that being touched an imaginative sense of shared humanity. (Diamond 1991: 55)

The terminally ill are not an exception to this rule either, just because they cannot any longer achieve many key forms of human *eudaimonia*. Perhaps it is true of them that they no longer have—to reuse Marquis’ phrase—“a future like ours”, and so will not achieve very much, or any, more *eudaimonia* before they die. Given that their futures have closed in on them in this radical way, you might call it “futile” to go on treating them with the full range of interpersonal hopes and expectations that you extend to any human, or to go on being as scrupulous about not killing them as you would be about anyone else. There again, you might also call it “respectful”; you might call it an expression of *pity*.

To pity someone in the sense I have in mind is to think, as MacIntyre puts it, “this could have been us”: this person too is part of human life and shares in human ideals, even if the idea of realising them in almost any way is quite hopeless for him/her. Such pity is a key part of our charity towards other people; of what is involved in recognising them as *people*, as fellow human beings.

(It is interestingly not something that follows in any very clear way from an Aristotelian virtue ethics where the focus is, eudaimonistically, on *maximal* forms of flourishing—on how much you as a being can develop, not on how many relatively undeveloped beings you can extend charity towards. Indeed it is a notorious fact about Aristotle that there is something rather ruthlessly bright-eye-and-gleaming-coat<sup>xxviii</sup> about his conception of *eudaimonia*, and that—not unconnectedly—he seems to have had little room in his ethics for the notion of pity.<sup>xxix</sup> Pity is, however, part of the sorts of virtue ethics that we may find in a Christian writer like Simone Weil. And that is the sort that I would defend, though I cannot defend it here.)

There are, as many people have emphasised, better and worse ways to die. And one of the worse ways to die is to have recognition of your humanity withdrawn

from you before it happens. Part of understanding why that is such a bad thing to have happen to you must, I think, depend on taking a different view from the criterialist's about the locus of the value of persons. For the criterialist, apparently, the value of persons consists in their instantiating the person-constituting properties. Aside from such instantiations, persons, at the most basic level, *have* no value (compare the familiar view in bioethics that being alive in itself is neither good nor bad); indeed there is a sense in which, aside from such instantiations, there *are* no persons. What such a view of the value of persons obscures from view is the notion of loving an individual in itself; it seems to be only the *properties* of an individual that a consistent criterialist can love—and then, only when they are duly manifested. But the idea that the object of love is always and necessarily some property-instantiation, common though it is, is a dogma. If property-instantiations can be objects of love (we may fairly ask), why not things in other metaphysical categories: individual human beings, for example?<sup>xxx</sup>

Alongside this dogma, perhaps there is another, related but slightly different, dogma at work in securing the foundations of criterialism. This is the dogma that whenever we answer the question “Why is X valuable?” by citing some valuable property V that X displays, what we must be really saying is always that it is (this instance of) V that is valuable, not X. But this too is a dogma, and unmasking it may be another way of setting ourselves free from the lures of criterialism. To cite V in explanation of the claim that X is valuable can, so to speak, be understood the other way around: you can take it that citing V is displaying a *symptom*, a *reminder*, of the value of X. When George Orwell, in a famous passage from his essay “A Hanging”, describes the growing fingernails and the refusal to walk through a puddle of the man who is about to be hanged, I take it that this latter thing is what he is doing.<sup>xxxi</sup> He is not saying that having growing fingernails is a value-giving property. Rather he is saying that *being human* is a value-giving property, the importance of which can be brought out for us by all sorts of exercises in the “assembling of reminders”—including such reminders as the growth of fingernails.

“But all the same, isn't there a profound difference between what we hope for *in general* for other people, and what we hope in the particular case of someone who is severely mentally handicapped?” There certainly is. There is always a distinction between the aspirations for other persons that are grounded in human *eudaimonia* in general, and the aspirations for specific other persons that are grounded in our particular knowledge of them as individuals. This distinction is just as applicable to the profoundly mentally handicapped as it is to anyone else. And it makes for the same commonalities, and the same differences, as with anyone else. If your friend is tone-deaf, there is no need to pretend that he will ever be Mozart; but that does not make it all right to take home all his CDs. If your friend loses both his legs in a skiing accident, you needn't spend the rest of his life pretending to him that he can still ski; but you miss something vital, too, if you bounce into his hospital ward straight after his double amputation with the words “So, you won't be wanting those ski-boots any more”. Many of the details of what counts as just and loving treatment of other persons are, certainly, set by the particular trajectories of their particular lives. But the most basic and elementary requirements of love and justice, is set by reference to the aspirations that arise from human *eudaimonia* in general. And this applies just as much to the mentally disabled as to anyone else.

In Nick Hornby's very funny novel about suicidal depression, *A Long Way Down*, Maureen, the mother of a severely mentally disabled child, says this:

This all began years ago, when I decided to decorate his bedroom. He was eight, and he still slept in a nursery—clowns on the curtains, bunny rabbits on the frieze round the wall, all the things I'd chosen when I was waiting for him and I didn't know what he was. And it was all peeling away... and I hadn't done anything about it because it made me think too much about all the things that weren't happening to him, all the ways he wasn't growing up. What was I going to replace the bunny rabbits with? He was eight, so perhaps trains and rocket ships and footballers were the right sort of thing for him—but of course, he didn't know what any of those things were... But there again, he didn't know what the rabbits were either, or the clowns. So what was I supposed to do? Everything was pretending, wasn't it? The only thing I could do that wasn't make-believe was paint the walls white, get a plain pair of curtains. That would be a way of [saying] that I knew he was a vegetable, a cabbage, and I wasn't trying to hide it. But then, where does it stop? Does that mean you can never buy him a T-shirt with a word on it, or a picture, because he'll never read, and he can't make sense of pictures? And who knows whether he even gets anything out of colours, or patterns? And it goes without saying that talking to him is ridiculous, and smiling at him, and kissing him on the head. Everything I do is pretending, so why not pretend properly?

In the end, I went for trains on the curtains, and your man from *Star Wars* on the lampshade. And soon after that, I started buying comics every now and then, just to see what a lad of his age might be reading and thinking about. And we started watching the Saturday morning television together, so I learned a little bit about pop singers he might like... thinking about these things helped me to see Matty, in a strange sort of way... I made up a son. (Nick Hornby, *A Long Way Down*, pp.118-9)

Of course there is—as Maureen herself recognises—something delusional, and something pathetic, in her treatment of Matty. For sure, her attitude to him is strained and exaggerated. But what we should not lose sight of is that it is a strained and exaggerated version of something morally indispensable: the kind of interpretive charity, the humanistic understanding and pity, that, I have argued here, we all need from everybody else all the time. In the words of an older writer: “Charity bears all things, believes all things, hopes all things, endures all things.”<sup>xxxiii</sup>

## Bibliography

- Bauby 1998: Jean-Dominique Bauby, *The Diving Bell and the Butterfly*. London: Fourth Estate, 1998.
- Carruthers 1992: Peter Carruthers, *The Animals Issue: Moral Theory in Practice*. Cambridge: Cambridge University Press.
- Chappell 1998: Timothy Chappell, *Understanding Human Goods*. Edinburgh: Edinburgh University Press.
- Chappell 2003: Timothy Chappell, "Absolutes and particulars", 95-117 in *Modern moral philosophy*, ed. A. O'Hear. Cambridge: Cambridge University Press.
- Davidson 1980: Donald Davidson, *Inquiries into truth and interpretation*. Oxford: Clarendon.
- Darwall 2006: Stephen Darwall, *The Second-Person Standpoint*. Cambridge, MA: Harvard University Press.
- Diamond 1991: Cora Diamond, „The importance of being human“, in David Cockburn, ed., *Human Beings*. Cambridge: Cambridge University Press.
- Engelhardt 1989: Tristram Engelhardt, "Ethical Issues in Aiding the Death of Young Children", in *Euthanasia: the moral issues*, edd. Robert M Baird and Stuart E Rosenbaum. Buffalo, N.Y. : Prometheus Books.
- Gaita 1991: Raimond Gaita, *Good and Evil: an absolute conception*. London: Routledge.
- Gaita 1998: Raimond Gaita, *A Common Humanity*. London: Routledge.
- Grau 2010: Christopher Grau, "Love and history", *Southern Journal of Philosophy* **48** (2010), 246-271.
- Grice 1957: H.P.Grice, "Meaning", *The Philosophical Review*, **66**: 377-388.
- Harris 1985: John Harris, *The Value of Life*. London: Routledge.
- Alison Hills, "Utilitarianism, contractualism and demandingness", *Philosophical Quarterly* **60** (April 2010), 225-242.
- Nick Hornby, *A Long Way Down*. London: Penguin, 2006.
- David Hume: *A Treatise of Human Nature* [1739].
- Kahane 2010: Guy Kahane, "Feeling pain for the very first time: the normative knowledge argument", *Philosophy and Phenomenological Research* **80.1** (2010), 20-49.
- Long and Sedley 1987: A.A. Long and D. Sedley, *The Hellenistic Philosophers*, Cambridge University Press, Cambridge, Vol.I, 67A
- Marquis 1997: Don Marquis, "An argument that abortion is wrong", in Hugh LaFollette, *Ethics in Practice*, First Edition (Oxford: Blackwell, 1997).
- MacIntyre 1999: Alasdair MacIntyre, *Dependent Rational Animals: why human beings need the virtues*. Chicago: Open Court, 1999.
- McMahan 2005: Jeff McMahan, "Our fellow creatures", *Journal of Ethics* **9** (2005), 353-380.
- Olson 2007: Eric Olson, Stanford Encyclopedia of Philosophy entry on "personal identity", <http://plato.stanford.edu/entries/identity-personal/>
- Parfit 1984: Derek Parfit, *Reasons and Persons*. Oxford: Oxford University Press.
- Patterson and Gordon 1993: "The Case for the Personhood of Gorillas", in Paola Cavalieri and Peter Singer (eds.), *The Great Ape Project* (New York: St. Martin's 1993), 58-77.
- Rachels 1990: James Rachels, *Created From Animals: The Moral Implications of Darwinism*. Oxford: Oxford University Press.
- Schapiro 1999: Tamar Schapiro, "What is a child?", *Ethics* **109.4**: 715-738.

- Singer 1983: Peter Singer, *The Expanding Circle*. Oxford: Oxford University Press.
- Singer 1991: Peter Singer, *A Companion to Ethics*. Oxford: Blackwell.
- Singer 1993: Peter Singer, *Practical Ethics*. Cambridge: Cambridge University Press.
- Singer 2002: Peter Singer, *Unsanctifying Human Life: essays on ethics*, ed. Helga Kuhse. Oxford: Blackwell.
- Sober 1994: Elliott Sober, *From a Biological Point of View*. Cambridge: Cambridge University Press.
- Strawson 1962: Peter Strawson, 'Freedom and Resentment', in his *Freedom and Resentment and other Essays*, London: Methuen, 1974.
- Teichman 1996: Jenny Teichman, *Social Ethics*. Oxford: Blackwell's.
- Tooley 1972: Michael Tooley, "Abortion and Infanticide", in Singer, ed., *Applied Ethics* (Oxford: OUP, 1987).
- Vlastos 1981: Gregory Vlastos, "The Individual as Object of Love in Plato", in his *Platonic Studies*, Princeton, NJ: Princeton University Press, 3–42, 2nd edn.
- Warren 1997: Mary Anne Warren, "On the Moral and Legal Status of Abortion", in H.Lafollette, ed., *Ethics in Practice* (Oxford: Blackwell, 1997).
- Williams 1985: Bernard Williams, *Ethics and the Limits of Philosophy*. London: Penguin.
- Williams 2006: Bernard Williams, *Philosophy as a humanistic discipline*. Cambridge: Cambridge University Press.
- Williams 2006a: Bernard Williams, "Hylomorphism", 218-230 in his *The Sense of The Past*. Cambridge: CUP.
- Ludwig Wittgenstein, *Philosophical Investigations*. Oxford: Blackwell, 1951.

## Notes

---

<sup>i</sup> Anscombe mistranslates *zur Seele* as “to a soul”; I have corrected this to “to *the* soul”.

<sup>ii</sup> Perhaps we could do ethics entirely or almost entirely without using the word or concept “person”: so numerous philosophers have at times suggested, including Jenny Teichman (1996), Bernard Williams (2006a), Christopher Coepe (2006), and myself (I have at least had this thought; I can’t now remember if it has ever seen the light of published day; something close to that thought is certainly at work in Chappell 1998, Chapter 4). However, I don’t pursue that idea here.

<sup>iii</sup> Criterialism, as I call it, is pretty much the same view as what McMahan 2005, following Rachels 1990, calls “moral individualism”. This is the view (Rachels 1990: 173) that “how an individual may be treated is determined, not by considering his group memberships, but by considering his own particular characteristics”.

<sup>iv</sup> Or something very close to criterialism. I need not commit myself yet on whether these authors are defending necessary-and-sufficient-conditions versions of a criterial view.

<sup>v</sup> Something like the criterial view pops up in many other places. For instance, it seems to be behind Olson 2007’s suggestion that “person” is a phase sortal: “Being a person may be only a temporary property of you, like being a philosopher”.

<sup>vi</sup> It might be suggested that (some instance of) Criterialism is a *bespoke* view of personhood, aimed only at defining what persons are for some specific purpose. Such a position is perfectly possible, of course, but I think it is very clearly *not* what is at issue in discussions like the four I cite here, all of which are aimed at completely general conclusions about “what persons are”: what they are in their essence, and for *any* possible purpose. (Thanks to an anonymous referee for discussion of this point.)

<sup>vii</sup> For more on personhood and language, see Patterson and Gordon 1993.

<sup>viii</sup> David Hume: “For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch *myself* at any time without a perception, and never can observe any thing but the perception. When my perceptions are removed for any time, as by sound sleep, so long am I insensible of *myself*, and may truly be said not to exist” (*Treatise* 1.6.3). Derek Parfit: “We could give a complete description of reality without claiming that persons exist” (Parfit 1984: 212). The Buddha: “There exists no individual, it is only a conventional name given to a set of elements” (quoted in Parfit 1984: 502).

<sup>ix</sup> Autism involves very various sorts of emotional dysfunction. It can involve *over*-emotionality and “inappropriate” emotional responses as well as a lack of emotional response. But that just underlines the absurdity of using “emotionality” as a necessary condition of personhood. For then autistics who are over-emotional will count as persons, autistics who are “cold” and “withdrawn” won’t.

<sup>x</sup> There is a surprising contrast between some recent philosophers’ willingness to grant moral standing to “potential persons” in the sense of metaphysically possible future human beings who admittedly do not exist right now, and their reluctance to do so to “potential persons” in the sense of young human beings who admittedly do exist right now.

<sup>xi</sup> “The wrongness of killing us is understood in terms of what killing does to us. Killing us imposes on us the misfortune of premature death. That misfortune underlies the wrongness... The misfortune of premature death consists in the loss to us of the future goods of consciousness... What makes my future valuable to me are those aspects of my future that I will (or would) value when I will (or would) experience them... What makes killing us wrong, in general, is that it deprives us of a future of value. Thus, killing someone is wrong, in general, when it deprives her of a future like ours.” (Marquis 1997: 95-96)

<sup>xii</sup> And it is nearly always *the* criterion, i.e. some *one* criterion, that we are talking about. (Warren’s list of six criteria is the only exception to this monism that I can think of.) The suspicion that the value of persons might be multiform not monistic, that there might be not just one simple and readily-articulable reason for valuing persons but *all sorts* of reasons, some of them far from simple or easy to put into words, is not at all a characteristic suspicion for criterialists.

<sup>xiii</sup> About the argument that I develop from this point on, about “how we treat” people, it is natural to ask: are these descriptive or normative claims? The answer to that, as quite often with Wittgensteinians like (I suppose) me, is: a bit of both. What is on offer is a phenomenology of (part of) the moral life. My claim is that the most basic and fundamental truths about this phenomenology makes certain options in thinking about ethics that might seem attractive and viable, such as criterialism, actually incoherent, because of its mismatch with what is basic and fundamental. Because of how things *are*, there are some things we *shouldn’t* think or do: both a normative and a descriptive thesis.

---

Another worry that might occur to some here: am I attacking criterialism because it is a *theory* of personhood? If so, isn't it hypocritical of me to go on to develop what is really just another theory of personhood? The answers are No and No. I'm not attacking criterialism because it's a *theory*, but because it's a *rotten* theory. A good theory (or if you dislike that word, a good philosophical account) of personhood will make good sense of our intuitions; this criterialism signally fails to do.

(My thanks to anonymous referees for pressing these questions.)

<sup>xiv</sup> The charming comparison is Singer's: Singer 1993: 89-90.

<sup>xv</sup> But when exactly? As part of the child's fourth birthday celebrations? On going to primary school? Would a Singerian society have a ritual for an induction into personhood?

<sup>xvi</sup> At least, the term "objective" is unfortunate if it implies, as I suspect it does in passages like the one from Strawson that I cite in the next Note, that there is something *non-objective* about participant reactive attitudes.

<sup>xvii</sup> "We look with an objective eye on the compulsive behaviour of the neurotic or the tiresome behaviour of a very young child, thinking in terms of treatment or training. But we sometimes look with something like the same eye on the behaviour of the normal and the mature. We have this resource [the objective attitude] and can sometimes use it. Being human we cannot, in the normal case, do this for very long, or altogether. But what is above all interesting is the tension there is, in us, between the participant attitude and the objective attitude. One is tempted to say: between our humanity and our intelligence. But to say this would be to distort both notions." (Strawson 1962: 66-67, with omissions)

<sup>xviii</sup> "One speaks and expresses emotions to and not simply at one's cat. Nonetheless, although we of course discipline them, we don't press claims against or hold our pets accountable in the same way we do with one another... Granted, when I look into my cat's eyes, I can't get over the feeling that he is looking also into mine in some personal way or shake the hope that he is seeing me in some way other than 'the guy who feeds me'. But I find it utterly impossible to sustain the thought that he can imaginatively enter into my point of view or acknowledge me as a being with an independent perspective." (Darwall 2006: 43)

Pets are an interesting case of something like a "pretend" interpersonal relationship, if "pretend" is the right thing to call it. We know that, unlike infants, our cats will *never* become responsive in the way that persons do. That doesn't stop us (or at least the dedicated pet-owners among us) from extending something interestingly like the constitutive prolepsis even to cats, hamsters, or stick-insects. Children, of course, do something similar with their dolls or other toys—and here "pretend" clearly is the right word. For a recent fashion to which similar comments apply see <http://www.nytimes.com/2009/07/26/magazine/26FOB-2DLove-t.html> . (Thanks to an anonymous referee for drawing this to my attention.)

<sup>xix</sup> "Intrinsic" in the sense of "essential", I mean; not in the sense of "non-relational", obviously.

<sup>xx</sup> Gaita 1991: 117: "An animal can suffer, but it cannot curse the day it was born. An animal can be afraid, but it cannot be ashamed of its fear and despite itself. An animal can be happy, but it cannot be joyous. An animal cannot give of its substance to certain pursuits and be admonished for doing so. One could go on almost indefinitely. The problems of life's meaning cannot arise for an animal. Only a being for whom life can be problematic can have a spiritual life, and therefore have a soul."

<sup>xxi</sup> For more examples applying this argument see Hills 2010: 235-7.

<sup>xxii</sup> So I say. In the case of trees, of course, there was at least one stern Victorian moralist who thought otherwise, namely W.E.Gladstone, whose hobby was chopping down trees.

<sup>xxiii</sup> For this speculation deployed for a different purpose, cp. Kahane 2010.

<sup>xxiv</sup> Cp. Gerald the gorilla in the well-known *Not the Nine O'Clock News* sketch:

<http://www.youtube.com/watch?v=MpbMm0433I>

<sup>xxv</sup> "Suppose that the experiments attempting to teach language to chimpanzees had been successful beyond their originators' wildest dreams. The apes in question gained a complete mastery of English within a few years, were able to attend school and later university, and made many close friendships with human beings... in these imaginary circumstances it would plainly be absurd to claim that the apes lacked moral standing, or had a moral importance that was lower than our own." (Carruthers 1992, Ch.3)

<sup>xxvi</sup> Carruthers' own view about animal rights is much closer to speciesism than to animal liberationism. He holds that there is an indirect argument, arising from contractualism, for granting all humans and no animals moral standing (for a brief exposition see his online paper "The animals issue" at <http://www.philosophy.umd.edu/Faculty/pcarruthers/The%20Animals%20Issue.pdf>). Since I think it is obviously true that (non-human) animals do have moral standing in their own right, based roughly on the wrongness of causing unnecessary suffering, damage, and destruction, it seems I am closer to animal liberationism than Carruthers is.

---

<sup>xxvii</sup> Jeff McMahan (2005: 366) writes that his “moral individualism” gives us reason to deny that it is a grave misfortune for a human individual who is disabled to have limited cognitive capacities. After all, it is not a misfortune for an individual animal to have capacities of roughly the same level; and according to moral individualism, it is only the individual creatures that we should be considering. So either both the human and the animal are unfortunate, or neither is.

This seems a very counter-intuitive consequence of “moral individualism”. Surely it matters how the creatures got into their states: at the very least, there must be a difference in fortune between a human individual who was born at a low cognitive level, and one who is at that level because of some horrendous accident? So far as I can see, McMahan must deny this. For presumably he thinks that my past stands to me as a relational characteristic, not an intrinsic one—and the whole point of “moral individualism” is that only “intrinsic characteristics” count morally.

<sup>xxviii</sup> Bernard Williams’ delightfully mordant phrase: Williams 1985: 46.

<sup>xxix</sup> The best presentation I know of the evidence for Aristotle’s blindness on the subject of pity is in Gaita 1991.

<sup>xxx</sup> For more on this point see Chappell 2003, Grau 2010, Vlastos 1981.

<sup>xxxi</sup> “It is curious, but till that moment I had never realized what it means to destroy a healthy, conscious man. When I saw the prisoner step aside to avoid the puddle, I saw the mystery, the unspeakable wrongness, of cutting a life short when it is in full tide. This man was not dying, he was alive just as we were alive. All the organs of his body were working—bowels digesting food, skin renewing itself, nails growing, tissues forming—all toiling away in solemn foolery. His nails would still be growing when he stood on the drop, when he was falling through the air with a tenth of a second to live. His eyes saw the yellow gravel and the grey walls, and his brain still remembered, foresaw, reasoned—reasoned even about puddles. He and we were a party of men walking together, seeing, hearing, feeling, understanding the same world; and in two minutes, with a sudden snap, one of us would be gone—one mind less, one world less.” (George Orwell, “A Hanging”, *Adelphi Magazine* August 1931)

<sup>xxxii</sup> My thanks for their comments and encouragement to a large number of people, including Julia Annas, Alex Barber, Chris Belshaw, David Benatar, Peter Cave, Stephen Clark, John Cottingham, Christopher Coope, Christopher Grau, John Haldane, Rosalind Hursthouse, Tim Mulgan, David Oderberg, Tim O’Hagan, Charles Taliaferro, Chris Tollefsen, and Garrath Williams; an audience in Leeds at the Society for Applied Philosophy in 2009; and three anonymous referees for this journal.



## **On the very idea of criteria for personhood**

### *Abstract*

I examine the familiar criterial view of personhood, according to which the possession of personal properties such as self-consciousness, emotionality, sentience, and so forth is necessary and sufficient for the status of a person. I argue that this view confuses criteria for personhood with parts of an ideal of personhood. In normal cases, we have already identified a creature as a person before we start looking for it to manifest the personal properties, indeed this pre-identification is part of what makes it possible for us to see and interpret the creature as a person in the first place. And that pre-identification typically runs on biological lines. Except in some interesting special or science-fiction cases, some of which I discuss, it is human animals that we identify as persons.

## On the very idea of criteria for personhood

Meine Einstellung zu ihm ist eine Einstellung zur Seele. Ich habe nicht die *Meinung*, dass er eine Seele hat.

My attitude towards him is an attitude towards the<sup>i</sup> soul. I am not of the *opinion* that he has a soul. (Ludwig Wittgenstein, *Philosophical Investigations*, II, p.152, tr. G.E.M.Anscombe)

Love is conditioned by its object, but love also yields its object. (Gaita 1991: 122)

### 1. Persons: humanism, speciesism, criterialism

Any ethical outlook much like ours will take as central some *primary moral constituency* (PMC): some class of creatures who all alike, and all equally, share in the highest level of moral rights and privileges. Most philosophical ethicists use “person” to mean at least “member of the primary moral constituency” (whatever else they may also mean by “person”). In this paper, that is how I shall use the word too.<sup>ii</sup>

So what makes any creature a person, a member of the primary moral constituency? One view is what I shall call humanism:

**Humanism:** Being human is sufficient for membership of the PMC.

Humanism faces two very common objections. The first is that humanism is arbitrary discrimination, like racism, sexism, ageism; humanism, in Richard Ryder’s famous phrase, is *speciesism*. The second objection, which is really just the same idea taken a little deeper, says that humanism bases its demarcation of persons on a biological property which is morally insignificant. Membership of this or that species—just like membership of this or that race, gender, or age-group—is, we are told, a morally insignificant detail. It is not the kind of thing that we should expect any important moral distinction to be based on.

In theory these objections could be put by anyone who rejected humanism. In practice they are usually put by proponents of another theory of personhood, which I shall call criterialism:

**Criterialism:** Actual possession of the criterial properties is necessary and sufficient for membership of the PMC.<sup>iii</sup>

Here are four very well-known enunciations of criterialism<sup>iv</sup>:

I propose to use “person”, in the sense of a rational and self-conscious being, to capture those elements of the popular sense of “human being” that are not covered by “member of the species *Homo sapiens*”. (Singer 1993: 87)

Persons are beings capable of valuing their own lives. (Harris 1985: 16-17)

An organism possesses a serious right to life only if it possesses the concept of a self as a continuing subject of experiences and other mental states, and believes that it is itself such an entity. (Tooley 1972: 82)

[The six key markers of personhood are] (1) sentience... (2) emotionality... (3) reason... (4) the capacity to communicate... (5) self-awareness... (6) moral agency. (Warren 1997: 83-84)<sup>v</sup>

As here represented, the criterial view is an all-purpose<sup>vi</sup> view about what personhood is, and why it matters, which is undeniably simple (as I shall argue, much too simple). Despite this simplicity, it is obvious, when criterialism is stated as baldly as it is in these four statements, that it has radical implications. Most noticeably, it implies a sharp contraction of the circle of intuitive moral concern. If rationality is a necessary condition of counting as a person, a member of the primary moral constituency, then all children under a certain age will fail to count as persons. (This bullet is explicitly bitten by Tristan Engelhardt: “If being a person is to be [*sic*] a responsible agent, a bearer of rights and duties, children are not persons in a strict sense” (Engelhardt 1989: 120).) If self-consciousness is a necessary condition, then many mentally handicapped humans will fail to count as persons. If the capacity to communicate is a necessary condition, then Jean-Dominique Bauby, of *The Diving Bell and the Butterfly* fame, would have failed to count as a person after his catastrophic stroke—had he not worked out how to signal with his eyelid.<sup>vii</sup> (Can whether Bauby is still a person after his stroke really depend on contingencies about whether he can still communicate?) If persons necessarily have “the concept of a self as a continuing subject”, and believe that they are such selves, then David Hume, Derek Parfit, and most Buddhists are not persons.<sup>viii</sup> If you cannot be a person unless you are capable of valuing your own life, then you cannot be both a person and a committed nihilist about value (i.e. someone who finds himself compelled to believe that no such thing as value exists). If persons have to possess “emotionality”, then Mr Spock in *Star Trek* is not a person, and neither, perhaps, are some extreme autistics.<sup>ix</sup> And so on. Criterialists can make personhood sound rather exclusive; it can seem as hard to qualify for personhood as it is to make membership of the Country Club.

Of course, these problems are consequences of taking these “markers of personhood” as indicators of necessary conditions for personhood. A less extreme version of the criterial view might take each of the markers on some list, or some or other combination of the markers, as a sufficient condition of personhood: any creature that displays just one of these properties, or some large-enough variety of them, will count as a person. We might then want to argue about what counts as displaying a property, or possessing a capacity. Must I actually evince emotions to pass Warren’s emotionality test? How often, and how convincingly? Do I possess the capacity to communicate only when I have learned a language? Or just when I have learned to sign, or to get others to read my thoughts and feelings? Or do I have the capacity to communicate all along, just in virtue of being a member of a species that communicates, linguistically and in other ways?

Here we touch on familiar and long-running debates about potentiality versus actuality in our assessments of personhood. It is commonplace for criterialists of Singer’s sort to point out that potential prime ministers obviously do not have the rights of actual prime ministers, so that “potential persons”<sup>x</sup>, or those with the

potential for what Don Marquis famously called “a future like ours”<sup>xi</sup>, can hardly be treated as the same moral category as actual persons, or “us”. It is equally commonplace for the opponents of Singer-style criterialism to object that there must be something wrong with an emphasis on *actual* properties that makes it a live question whether one is a person while asleep or under general anaesthetic.

Clearly there are problems about at least some of the suggestions listed above about what the criteria of personhood might actually be. But not all those suggestions are implausible, and it would be easy enough, in principle, to weed out or refine the implausible ones. True, it is striking how little criterialists feel they need to do this; in practice, criterialists often seem perfectly happy to stick with what, intuitively, seem implausibly over-demanding criteria for personhood.

However, the plausibility or otherwise of the various suggested criteria is not my main concern here. Obviously enough, other criteria could be devised which did not have these consequences. The key difficulties for criterialism that I want to emphasise in this paper are three.

The first, which I can state very quickly, is about the argumentative method underlying criterialism. To put it at its simplest, it *looks* as if criterialists typically adopt their preferred criterion of moral significance because they think it explains moral intuitions which we have anyway. The criterion is adopted, in other words, because it doesn't have counter-intuitive consequences.<sup>xii</sup> But then, the same criterion is deployed, by the very same authors—Peter Singer and John Harris, for example—precisely to *produce* counter-intuitive consequences, such as the moral acceptability of killing small babies or the mentally disabled. There is, I think, a serious question as to what can possibly be going on here. It is not entirely obvious why building a moral theory out of one intuition or set of intuitions that we certainly have, and then using this theory to attack other intuitions that we also certainly have, is anything more than arbitrary selectiveness. Perhaps this is a question that all system-building moral theory faces.

The second difficulty for criterialism, which I shall also mention only very briefly, is that criterialism can seem to imply a quite implausible picture of moral reasoning. Are we really to suppose that what we do when we come across any being we haven't met before is *first* see whether it satisfies our criterion for personhood, and *then*, if it does, start treating it as a person? One is reminded of that well-known caricature the Cartesian Detective: we start with the hypothesis—the *opinion*, as Wittgenstein says in my epigraph—that the new being is a person, and feel justified in acting in line with that hypothesis, just insofar as the behavioural evidence confirms it. If criterialism is to be at all plausible, it cannot in this way be a general view about our typical real-time interaction with others. Such a picture might apply to some very exceptional cases, some of which we will discuss in section 4b. But in general, criterialism will have to be a view about the *criterion of rightness* for taking other creatures to be persons, not about *our decision procedure* for assuming that they are persons. For quite obviously, our normal interaction with others typically involves no such Cartesian-Detective rigmarole, and it takes a good deal of philosophical callousing of the soul not to find something absurd, and morally repugnant too, about the idea that it even could. But if in general criterialism only states the criterion of rightness, and not our decision procedure, for taking other creatures to be persons,

then it still faces a further objection, my third—which it is the main purpose of this paper to develop.

This third objection to criterialism, on which I shall spend the rest of this paper, is connected to this second point. My claim is that the sort of properties that criterialists home in on are not criteria of personhood at all. Rather, they are dimensions of interpretation of beings that we *already* take to be persons.

For my own part, I have always been aware of at least a vague sense of discomfort and unease here, the moment we start talking about these (or any other) properties as if they were criteria for personhood of the sort the criterialists have in mind. There is a hard-to-articulate feeling that, despite the considerable first-blush plausibility of criterialism, it leaves out something crucial; a feeling of the kind that is naturally expressed in ordinary conversation by a trailing-off “Yes, *but...*”. I suspect that my unease is not unrelated to the unease that Raimond Gaita is bringing out when he writes (Gaita 1991: 115) that it is

naïve to think that the kinds of properties which interest philosophers when they ask what is a person, and which we often share to some degree with higher animals, play the kind of role in our treatment of one another which is assumed by those who argue that differences in treatment must be justified by relevantly different properties of those kinds. That philosophical perspective from which we are encouraged to reassess our sense of how we might justifiably treat animals distorts and indeed cheapens our understanding of human life.

Gaita 1991 is a marvellous exploration of one philosopher’s attempt to make articulate that trailing-off “Yes, *but...*”. This essay offers another philosopher’s. My attempt is not necessarily the same as Gaita’s, but not necessarily in competition with it either.

And it is perhaps worth saying before I begin that what I am trying to do here, what Gaita was trying to do, is *hard*: hard to articulate, hard to state clearly, hard to see in the round, hard to apply. Again and again criterialism wins out in philosophers’ discussions of personhood because, conversely, it is dead easy; it is such a simple and straightforward view of personhood. There are plenty of spurious analogies between science and ethics around, but at any rate we can trust this one: no more in the case of the nature of the person than in the case of particle physics does the fact that a view is *simple* in any way improve its chances of being *true*.

## **2. The proleptic view of personhood**

Contrary to what criterialism seems to suggest, we do not look for sentience or rationality or self-awareness in a creature as a test to decide whether or not that creature counts as a person. It’s the other way round. Having once decided, on other grounds, that a creature *is* a person, we know that this makes it the kind of creature which is likely to display sentience, rationality, self-awareness and the rest of the personal properties. Hence we look for displays of these properties from the creature. That is to say, we treat it as a person *in advance* of any such displays.<sup>xiii</sup>

The biggest and clearest example of this is an activity that is absolutely central to human life: parenting. Parents are, of course, aware of the differences in rationality, linguistic capacity, self-representing ability, and so forth between young human beings at various ages and developmental stages. Nonetheless, a parent's attitude towards her child is always, basically, what Wittgenstein famously calls "an attitude towards a soul". She does not do what criterialism (especially if understood in decision-procedure terms) might seem to imply she should do: start out by treating her child as an inanimate object, like a sofa or a fridge or a rubber-plant, and grudgingly consent to adjust her attitude to it, one little step at a time, only as and when it proves itself more than inanimate by passing a succession of behavioural tests for at least having interests like a snail's perhaps,<sup>xiv</sup> and then later on for the interests of some more advanced kind of animal such as a puppy, and eventually for full criterialist personhood.<sup>xv</sup> If parents *did* treat their children in this almost behaviouristic fashion, the parents would be callous monsters, and the children would be basket-cases. In a world where parents generally accepted this sort of criterial view, and applied it directly in their parenting practice, even the best-adjusted of us would be a wolf-child.

Fortunately, however, parents are not criterialists, either in their decision procedure or in the criterion of rightness for ascriptions of personhood that they actually employ. Rather, a parent treats her child from the very beginning—and from before it is literally and actually true—as a creature that can reason, respond, reflect, feel, laugh, think about itself as a person, think about others as persons too, and do everything else that persons characteristically do. From the beginning her attitudes towards the child are not only "objective" (to use Strawson 1962: 67's slightly unfortunate term<sup>xvi</sup>): they also include what he famously calls "participant reactive attitudes", of just the same sort as she adopts towards anyone else.

Of course, as Strawson implies<sup>xvii</sup>, these participant reactive attitudes' scope may be more restricted in the case of a very small baby. The point is that they are there; and that like the baby—often indeed in advance of the baby—they *grow*. As Alasdair MacIntyre notes in a recent discussion of the development of practical reasoning in the individual person (1999: 90), ethically and psychologically adequate parenthood centrally involves "a systematic *refusal* to treat the child in a way that is proportionate to its qualities and aptitudes" (that is, its actual and non-idealised qualities and aptitudes, the kind of qualities and aptitudes that interest the criterialist).

The parent who says to her three day old baby, "What do you think? Should we give you some more milk now?", or to her three year old toddler "How kind of you to share your toys with your sister when she's sad!", is not deluded about a neonate's capacity for interpersonal practical reasoning, or a toddler's capacity for calculatedly empathetic altruism. But she is not making a sentimental joke either. She is treating the baby or the toddler proleptically, in the light of the ideal of personhood. By years of treating her children as creatures who "have the personal properties"—in the sense that interests the criterialist—she makes it true that they *are* creatures who have the personal properties in just that sense.

The child staggers across and plonks a book down on his sibling's lap; the parent's reaction is "How kind of you to let her share your book!" Is that, in fact, what the toddler was doing? The right answer to this question can be: "Yes it is—once the

parent has given this reaction.” The parent’s reaction is an interpretation of the child’s deed. Before the reaction, perhaps, there was no fact of the matter about what the little boy was doing: the child did not know himself what his action was. For all he knew about it, his action might as well have been simply a random sequence of bodily movements. But now that his mother offers her interpretation, the child, on his mother’s authority, learns to see himself a certain way. Because his mother frames his act as one with a certain meaning (as the giving of a gift), and because the child sees his mother as seeing his act this way, and because his mother sees the child as seeing her as seeing his act this way... (Grice 1957), the act *comes to have* that meaning.

Paul Grice’s famous work on meaning and intention is even more relevant to another clear example of this sort of prolepsis, namely language-learning. A baby begins by babbling—that is, by producing all the sounds the human mouthparts can make. Some of these sounds get a response because they are sounds that occur in the parents’ language, while others don’t because they are not. (The inability of many English adults to pronounce the Scots *ch* in *loch* or the Welsh *ll* in *Llanfair* is a learnt inability, not mirrored in English babies, who can and do make both sounds until conditioned not to by English adults.) The baby “homes in” on the sounds, and then on the patterns or combinations of sounds, that get a response. Then it learns to correlate particular patterns of sounds with particular contexts by the same homing-in process. What the parents say is “She’s learned to say ‘tiger’” (my own first child’s first word). What has happened is that the parents, by treating their daughter *as if she has* the word for, and then the concept of, a tiger, have made it true that she *has* that word. (And, a little later perhaps, that concept too; though I am not suggesting that any one word or concept can be learned on its own: in Wittgenstein’s familiar image, “light dawns gradually over the whole” (*On Certainty* 141).)

Stephen Darwall too talks of prolepsis when describing a third connected process, namely the process whereby we come to be, and to hold others, morally responsible (Darwall 2006: 87-88):

There seem to be many cases where we wish to hold others accountable though we seem to have very good evidence that they are not free to act on moral reasons in the way our practices of holding someone fully responsible seem to presuppose... In some instances, for example with children, we seem simultaneously to move on two tracks in the process of inducting them into full second-person responsibility, sometimes treating them proleptically as though they were apt for second-personal address as a way of developing moral competence while nonetheless realising... that [at their developmental stage] this is an illusion.

Or an illusion, at any rate, if you agree with the criterialist about what it takes to be “apt for second-personal address”, and/or about what it takes to “have the personal properties”. My point, with which Darwall might or might not agree—I’m not sure—is that what it takes does not have to be already-realised and –actualised capacities of the sorts that interest criterialists. (The mother who asks her neonate child “Shall I give you some milk?” is certainly engaging in second-person address, *despite* the manifold incapacities, perfectly well-known to her, of neonates.)

There is at least this much truth in Dennett 1987's well-known idea of the "intentional stance": persons constitute each other as persons, both as agents and as patients, by *treating* each other as persons, as agents and patients. Something like the Davidsonian "principle of charity" (Davidson 1980) is at work in our mutual interpretations. By charitably, and proleptically, interpreting the other as a person, I *make* him a person.

The looseness and ambiguity of the term "person" is evident when we say things like this. In ordinary language, "person" can certainly, at one end of its semantic range, mean something like what Singer and other criterialists want it to mean, a "rational and self-conscious being" (or whatever). Yet at the other end of its semantic range, the ordinary-language term can also be a synonym for "human animal". It seems to be part of the *point* of our person-concept to allow those young humans who are persons only in the "human animal" sense to enter into a continuum of meaning towards the far end of which they can become persons in much stronger senses.

Many things in our shared life can be seen in this sort of way as aspirations towards an ideal which, like any ideal, is never fully realised. The basic sort of idealisation, the sort that grounds attributions of personhood, is not the only sort. We ascribe the rights to freedom of assembly and freedom of speech to every human being, even though we know that small babies can't speak and don't assemble (not without help), and that acutely asocial loners may well never (or at least never again) actually exercise either right. The behaviour of parents, besides (as described) involving the person-constituting prolepsis in a particularly clear form, is also full of other similarly-patterned proleptic idealisations that might, on a criterialist's view, be regarded as quite unrealistic. Parents give their children the opportunities to play the cello, to learn Spanish, to use their monthly allowances sensibly, to be polite to Great Aunt Maud (...), even though the parents may be fully aware that every one of these opportunities will in all probability be spurned. In minimally decent jurisdictions the treatment even of hardened criminals displays the same pattern of basic *openness* to the convict, however much of an unrealistic idealisation this openness may seem to involve: the jurisdiction goes on offering the criminals opportunities to reform, even though it is as good as certain that these opportunities will not be taken up. (I think that something like this refusal to "close the door" finally and irrevocably on the criminal is also, in connection perhaps with some thoughts about the state's obligation to be fundamentally benevolent to the citizens for whose sake alone it after all exists, the ground of a good reason to reject capital punishment. But that's another story.)

In general, the idealisation that accords X the status due to a  $\phi$ er does not have to involve any kind of expectation that X will ever actually  $\phi$ , or even can  $\phi$ . What it *does* have to involve is the understanding that X is of a kind that characteristically includes  $\phi$ ers. The human kind includes individuals who characteristically (at the right point in the human life-cycle) will freely assemble and speak. That is why it is not ridiculous to give the rights to free assembly, speech, and political action, e.g. becoming a senator, to any and every human, including asocial misanthropes, those (as yet, or as of recently) physically incapable of speech, and those who choose never to utter a word. Contrast the equine kind, which does not characteristically include any such individuals. That is why it *is* ridiculous to give these rights (or the status of a senator) to any horse.



Considered as a kind, children characteristically include individuals who learn the cello or Spanish or wise use of money or good manners to venerable relatives (well... eventually). That is why the parents who try to teach these lessons to their children are not acting absurdly even if they know in advance that *this particular* child is not going to respond—whereas parents who tried to teach these lessons to the family goldfish or the rubber plant *would* be acting absurdly. Similarly, there is nothing absurd about offering the old lag a place on the rehabilitation programme, even if it seems certain that *this* old lag will always refuse or abuse the offer: and the reason why not has to do with the general fact that some old lags do sometimes reform.

*A fortiori*, there is nothing absurd about giving the rights of free speech, assembly, private choice and the rest of it to women, or to other races, or to homosexuals or transgender people, or to the disabled, or to gingers, or to the various other sorts of historically marginalised groups. People in these categories are not another kind, as horses and hamsters are. Their well-being is *human* well-being, just as white male heterosexuals' is, and just as horses' and hamsters' well-being is not. That, in a nutshell, is why speciesism is *not* like racism or sexism.

Clearly, then, this is not just a point about the induction of babies and very small children into the social life of humanity at large. All humans, even the healthy intelligent mature well-adjusted independent affluent adults that typical criterialism is so clearly focused on—the Country Club again—are always only incompletely and impurely agents. Incompletely, because we are never all agents can be. However well things go, we never do or even can reach the full potential of human well-being, which would include a fully worked-out articulation of our own agency, a full coordination and marshalling of our own forces and powers for action in the world: towards this we are always only, at best, on the way. And impurely, because each of us always *can* be interpreted, by a sufficiently determined sceptic, as not really an agent at all, but a victim of the instinctive or other determining external forces that constantly threaten anyone's agency. To extend “an attitude towards a soul” towards *anyone* is always, and not just with children, an idealisation. As Tamar Schapiro notes within a Kantian framework (Schapiro 1999: 723):

[T]here is a sense in which no one, regardless of age or maturity, is able to achieve autonomy on Kant's view. This is because the notion of autonomy in Kant is an ideal concept which outstrips all possible realisations in experience. Strictly speaking, every instance of human willing is necessarily an imperfect realisation of transcendental freedom, and every virtuous character necessarily falls short of perfect virtue. And yet the applicability of the moral law depends on our mapping these ideal concepts onto ourselves and one another for the purposes of guiding action. So we are to regard the social world as a community of autonomous agents despite the fact that perfect realisations of autonomy are nowhere to be found.

The criterialists are right to insist that properties like self-awareness, emotionality, rationality, and so forth are crucial parts of our concept of a person. It does not follow that these properties can be used as the criterialists want to use them: as *tests* for personhood. Even at the cool-hour level of the criterion of rightness—never mind the

real-time level of deliberative procedure—to treat someone as a person is not to put a tick in the box by her name, to show that she has passed some inspection or met some standard, of rationality or self-awareness or emotionality or whatever. Indeed, it seems no less arbitrary discrimination to say “Sorry, you’re not rational or self-aware or linguistically capable or emotionally responsive enough to count as a person” than to say “Sorry, you’re not white enough to count as a person”. Behavioural properties like rationality, self-awareness, emotionality are not tests for, but parts of the ideal of, personhood. To treat someone as a person is to engage with him as the kind of creature to which that ideal applies. So to treat him is not, at the deepest level, a response to his *behaviour* at all, but to his nature. To see some creature as a person is to take an attitude to that creature which, *before* any behavioural evidence comes in, is already different from our attitudes to creatures that (we think) aren’t persons. This is the second-person standpoint (to use the title of Darwall’s outstanding book): the attitude we take to persons, and do not take to sofas, fridges, or rubber plants.

The fact that it takes charity, in more than one sense, to constitute persons imposes on persons a crucial sort of vulnerability and dependence: “The personal reactive attitudes rest on, and reflect, an expectation of, and demand for, the manifestation of a certain degree of good will or regard on the part of other human beings towards ourselves” (Strawson 1962: 70). All human persons are, in Alasdair MacIntyre’s splendidly apt phrase, “dependent rational animals”, as MacIntyre himself emphasises (MacIntyre 1999: 81-85):

In most moral philosophy the starting point is one that already presupposes the existence of mature independent practical reasoners whose social relationships are the relationships of the adult world. Childhood, if noticed at all, is a topic that receives only brief and incidental attention... To become an effective independent practical reasoner is an achievement, but it is always one to which others have made an essential contribution. The earliest of these relate directly to our animal existence, to what we share in our development with members of other intelligent species... [But what] we [also] need from others, if we are not only to exercise our initial animal capacities, but also to develop the capacities of independent practical reasoners, are those relationships necessary for fostering the ability to evaluate, modify, or reject our own practical judgements... Acknowledgement of dependence is the key to independence.

We are vulnerable to each other, and dependent on each other, because it is remarkably easy to attack persons by withdrawing the second-personal attitude from them. Part of the reason why it is so easy to stop seeing others as persons is, as I said before, because the interpersonal attitude always involves a degree of idealisation. There is always *some* truth in sceptical or cynical views that reject that idealisation. (For example, the human person *is* a physical object—even if it is not “merely” a physical object.)

We fail to see others as persons in a small way when we explain away what others do, or refuse to listen to them or to take their wishes and ambitions seriously. (Parents are never perfect in charity; children’s protests against such treatment are a recurring theme of family life.) We do it in bigger ways when we “use” others, or (in Kant’s phrase) fail to treat them as “ends in themselves”. At the limit, we can withdraw the interpersonal attitude from others completely, by pretending that they

simply aren't there, as British and Irish settlers did to the aboriginal Australians under their notorious legal doctrine of *terra nullius* (see Gaita 1998). Or we can do it by treating them merely as physical objects, as often happens in pornography, violence, and murder. Or, lastly, we can do it by declaring them "non-persons", as the Nazis did to the Jews—and as criterialists routinely do to the very old, the very young, and the very disabled.

The personal properties, then, set the content of the ideal of personhood. They do not set the boundaries of the class of persons (the PMC, as I've also called it). What does set the boundaries? That is my question in section 3. My answer will be the humanist answer already advertised in section 1: roughly, the boundaries of the class of persons are the boundaries of the human species.

### 3. The truth in speciesism

Rubber plants, fridges, sofas, human babies: we come to the last of these four sorts of things with a very different attitude from our attitudes to the other three sorts. And my point is that we *come to them* with different attitudes. Our attitudes to them are not responses to evidence that we gather *after* we have "come to them". *The moment I meet* a rubber plant or a fridge, I am perfectly sure, unless my sanity is in question, that it will not attempt to talk to me or tickle me—as, say, a baby might.

What makes me so sure of that? Well, simply the sort of thing a rubber plant or a fridge is, and the kinds of behaviour that it is natural to expect from that sort of thing. Rubber plants and fridges are not the kind of thing that tries to talk to you or tickle you. Babies are. And if you ask "What kind of thing is that?", the obvious answer is: "Young human beings". If some creature is a human being, it is completely natural to expect it to be the kind of thing that might (at least if it's a family member) try to talk to you, or tickle you, or otherwise demonstrate its (budding) mastery of the kinds of mental capacity that criterialists take to be criteria. (Talking to someone and tickling someone both require interesting interpersonal capabilities. To do either, you have to want things to be a certain way for someone else—roughly, you have to want to get them to think a thought, or to feel a tickle. These are wants with quite sophisticated contents.)

This sort of expectation is based squarely on the nature of the creature in question. And that means, on its nature *as a member of the human species*. Our understanding of which creatures are persons, and so may be expected to exhibit the personal properties, is not formed on an individual basis but on the basis of generalisation from experience of humans in general. I have already mentioned Wittgenstein's remark that "my attitude towards him is an attitude towards the soul"; on the same page of *Philosophical Investigations*, Wittgenstein also says that "the human body is the best picture of the human soul" (*der menschliche Körper ist das beste Bild der menschlichen Seele*). And perhaps what he means is this: that when we see that some creature is a human, just seeing that is enough to license us to adopt the interpersonal attitude to that creature.

We know in advance how we can expect any particular rubber plant or fridge to behave, because we can draw on a bank of past experience (our own and others') of

rubber plants and fridges. In this sense, the fact that a thing is a rubber plant or a fridge sets the scope of our expectations about it. In just the same way, we know what kinds of capacities and properties humans in general have. It is that background knowledge that we draw on, whenever we apply the interpersonal attitude to any individual human being, simply on the basis that it *is* a human being.

We obviously do the same with other species of animals. When I meet a cat or a gerbil, I treat it neither as an inanimate object like a sofa or a fridge or a rubber plant, nor as a person. (Not, of course, that I treat all inanimate objects the same, either. Sofas, fridges, and rubber plants are very different things, especially rubber plants.) You can play games with a gerbil. With a dog or a cat, you can even have a cuddle (in the reciprocal sense, I mean).<sup>xviii</sup> Neither is possible with a fridge. On the other hand, you would be wrong to expect a dog to play a good hand of contract bridge, or a cat to give you good investment advice, or a gerbil to read your mind, as a human being might. What you expect of an individual dog or individual cat will be different from what you expect of an individual gerbil, and both will be different from your expectations of inanimate objects. In each case, your expectations will be set by your background knowledge of the different kinds in question. Where dogs, gerbils, cats, and indeed all other kinds of living thing are concerned, this obviously means the different *species* in question. Why should it be otherwise with humans?

These remarks should enable us to see—*pace* the criterialists—how central it is, to our moral responses to the things and creatures around us, to assign them to species. It is species-assignments that enable us to answer what you might reasonably think is the most basic question of all about any thing that might be in front of us at any time: namely, what kind of thing it is. Until we know what species it is, we do not know how it may act, or how to treat it: what is good for it, how we might harm or hinder it, what counts as health for it, and so on. The nature of the creature determines what well-being is for it; and its nature is its species. To put it another way, the best way to answer “What is it?”, when this question is asked about any individual animal, is to give the name of its species. Or as we might ask moral individualists like McMahan: “If species-membership isn’t an intrinsic property, what *is*?” Singer, Rachels, and McMahan never tell us why *being human* cannot count as any creature’s “own particular characteristic”. It can’t be merely because *being human* is a group membership (as McMahan 2005 suggests). For there is nothing to stop a given property from being both a “group membership” and an “own particular characteristic”; that is how it is with *being a vertebrate*, for instance. To put it more formally, nothing stops a given property from being both a relational property, and also an intrinsic property.<sup>xix</sup> Indeed, if Kripke is right in his well-known thesis of the metaphysical necessity of origin, such properties must be very common.

This shows how humanism can be true without falling prey to a charge of arbitrary speciesism. To treat any species of creatures as persons is to treat them, not as successful passers of some test for the personal qualities, but on the assumption that they are individuals of a species the ideal for which is (roughly) the highest attainable development of those qualities. In this sense, references to species are an ineliminable part of our talk about persons.

In that sense, but not in some others. We are now in a position to respond to four familiar objections to “speciesism”. I consider these in section 4. The aim of my

response is not of course to defend speciesism, if that is defined as “*arbitrary* moral discrimination on the grounds of species”. But it is to defend humanism, the view that not all moral discrimination on the grounds of species *is* arbitrary.

#### 4. Four objections to speciesism

*a. All animals are equal.* Peter Singer (1993: 57) writes that “No matter what the nature of the being, the principle of equality requires that the suffering be counted equally with the like suffering of any other being”. Nothing I have said so far commits me to rejecting this application of the principle of equality. If I were to accept it, I would want to add, first, that plenty of other things besides suffering “count”, and secondly, that however alike human and animal *pain* (the physical sensation) may be, human and animal *suffering* (the state of mind) are mostly quite unlike. A cat can be tortured like Gloucester, but it cannot suffer like King Lear.<sup>xx</sup> And suffering matters more than pain.

But perhaps we should reject this application of equality anyway. Right now there is (quite likely) monstrous pain, and a fair amount of suffering too, on the plains of the Serengeti, where a wounded buffalo is being eaten alive by a pride of lions. Singer’s principle of equality requires us to say that our duty to prevent the buffalo’s pain and suffering in the teeth of the lions is directly proportional to the amount of pain and suffering involved. If any implication of any moral theory is absurd, this one is. There just isn’t any duty, not even a *prima facie* one, for us to act as Zoological Pain Police. (Don’t say: “We can’t intervene, because the lions might eat *us*”. Where is your moral courage? And don’t say: “But it causes less pain in nature overall if we don’t intervene in nature in this sort of way.” As many writers on the theological problem of evil are keen to remind us, most animal suffering is gratuitous. If lions can be caught and trained to jump through hoops, perhaps they can be caught and trained to kill their prey cleanly before eating it. So is anyone *trying* to train them? Of course they’re not.) The vast majority of the huge amount of animal pain and suffering that happens every day is simply none of our business. “It is not an accident or a limitation or a prejudice that we cannot care equally about all the suffering in the world: it is a condition of our existence and our sanity” (Williams 2006: 147).<sup>xxi</sup>

So it looks like Singer’s principle that all suffering counts for the same, no matter what the species involved, is just false. On the other hand, some weaker principle of this general sort seems clearly true: perhaps “Animal suffering is bad for the same sort of reasons as human suffering is bad”, or something like that. What also seems clearly true is that humans should do what they can not to generate animal suffering, and to prevent animals from being needlessly hurt or damaged *by humans* (and sometimes, though less often, by accident or by each other). This is a much weaker principle than Singer’s equality-of-suffering principle. But perhaps even this is strong enough to generate a good argument for vegetarianism. I suspect it all depends what counts as needless hurt or harm.

Hence “the limit of sentience” is not “the only defensible boundary of concern for the interests of others” (Singer 1993: 58). Certainly that limit marks one distinction between the ways we should take an interest in the welfare of other creatures. For example, hacking a frog to pieces is wrong in a way that hacking an

orange to pieces is not, partly because an orange certainly cannot feel pain, and a frog pretty definitely can. But only partly. Quite aside from the pain involved, destroying a living creature for fun is wrong in a way that destroying a piece of fruit for fun is not. (Hacking trees, or anaesthetised frogs, to pieces for fun is usually wrong too.<sup>xxii</sup>) Unless we think that nothing matters except pain and pleasure—and this is not, e.g., Singer’s position either (Singer 1993: xi)—there is no reason to think that the distinction between sentience and non-sentience is the only fundamental moral distinction we can make among beings; nor that it is the most important.

There are some interesting complications about the notion of sentience. If “sentience” means simply “capacity to feel (sensational) pain and pleasure”, then sentience does not fit the story that I have developed about how the personal properties come into being. The capacity to feel pain- and pleasure-sensations is certainly one of the capacities that we expect to find in any human person. But it is not, unlike rationality or intentionality, a capacity whose presence in any person is even partly constituted by others’ idealising assumption, expectation, or hope that it is there. Whether or not a being can feel pain does not depend on others’ adopting “participant attitudes” or taking the “intentional stance” towards him, in the way that (I have argued) his capacities for self-consciousness, rationality, and suffering do. I don’t think this shows that sentience is a counter-example to my thesis about how the person-constituting properties are developed. Rather, I think it shows that sentience is not one of the person-constituting properties. A creature which had the (other) person-constituting properties, but lacked the capacity for pain- and pleasure-sensations, would still undeniably be a person.<sup>xxiii</sup> A creature which had the capacity for pain and pleasure, but lacked the (other) person-constituting properties, surely would not be.

*b. Aliens, talking chimps, Locke’s parrot, and a possessed fridge.* The position that I am defending under the name of humanism says that, normally, to be a person it is *sufficient* to be a human being. Not *necessary*, because humanism does not imply that members of other species could not be drawn into the moral community of human persons. On the contrary, I can see at least three distinct ways in which this could happen. One is that we could discover a whole species—aliens, or angels, or what have you—which is like the human species in that its members are characteristically persons (in one sense or another). A second is that a species already known to us might, *en masse*, somehow change so as to become characteristically persons. Or, third, some particular individual creature—Locke’s parrot, say (Locke, *Essay* 2.27), or Washoe the chimp (Singer 1993: 111)<sup>xxiv</sup>, or some group of creatures, like Peter Carruthers’ Chimp Class of ’92<sup>xxv</sup>—might start to count as persons, perhaps e.g. by displaying the personal properties, even though this behaviour is otherwise completely untypical of the species in question. (Come to that, I suppose a possessed fridge might, in principle, start displaying the personal properties.)

In these special cases, we perhaps *will* have to appeal to something like the criterial view of personhood as our decision procedure for the vexed cases. Apparently we will have to ask whether this alien species, or this new class of highly-educated chimps, or this individual chimp, parrot, or spooky fridge, displays something like the criterialist’s personal properties, *before* we can decide whether to count these beings as persons.

Doesn't this show that my arguments against criterialism were mistaken? Not at all. It does show that something like criterialism may perhaps be our decision procedure if and when we encounter a being *that we do not already know how to classify as a person or not a person by way of its species*. (Perhaps; actually I doubt even this much is true, but at any rate it is the most that is true.) However, it would be utterly misleading to generalise from thought-experiments about these special and rare cases—almost all of which, to date, are imaginary—to alleged conclusions about the normal cases. (It would be as misleading to do that as it would be to generalise the justificatory/ rational structure of *every* belief we have about the external world from the justificatory/ rational structure of the external-world beliefs that we might form when we are thinking about the classic, and nearly always imaginary, sceptical predicaments.) Unfortunately, just this generalisation is a criterialist stock-in-trade. The most famous example of it is perhaps Mary Anne Warren's derivation of the personal properties (her list was quoted in section 1) from an imagined meeting with aliens. But even if Warren is right that this is the decision procedure that we would have to adopt with *aliens* to settle whether they were persons in some extraordinary imaginary scenario, nothing at all follows about our beliefs (either at the decision-procedure or at the criterion-of-rightness level) about whether any *humans* are persons in a multitude of ordinary real scenarios. For we have met humans before.

But anyway, I suspect Warren's description is much too black-and-white to get anywhere near characterising the decision procedure that would really be involved in such a meeting. (This is why I say, above, only that we would probably have to use *something like* the criterialist approach in alien encounters and similar.) For one thing, there is not the slightest reason to think that, in such a meeting, *yes* and *no* would be the only possible answers available to the question "Are these persons?". Even *Star Trek* yields a rather subtler answer—"It's life, Jim, but not as we know it"; the world, after all, is a complicated place, and it constantly throws us surprises. For another, our responses in such situations would be—or should I say *will* be?—based on all sorts of factors and reactions, most of them probably defying explicit articulation; on judgements about "forms of life", in fact. It would not and could not be based on some neatly finite check-list of diagnostic properties like Warren's. Looking beyond *Star Trek*, science fiction is full of wonderfully complex and subtle explorations of such encounters. The contrast with the mechanical simplicities of a criterialist list such as Warren suggests is instructive—and not *entirely* to be dismissed as unfair on the grounds that Warren is, after all, not writing science fiction.

*c. The two-species scenario.* A third interesting objection to speciesism is offered by Peter Carruthers:

Suppose it had been discovered that human beings in fact consist of two distinct species, otherwise hardly distinguishable from one another, the members of which cannot inter-breed. In these circumstances it would plainly be objectionable for the members of the majority species to attempt to withhold moral rights from the members of the minority, on the mere ground of difference of species. This, too, would be obvious speciesism. (Carruthers 1992 Ch 3)

Carruthers, I think, is quite right to claim that it would alter nothing morally speaking if we discovered that "the human species" was in fact two different

biological species—perhaps *homo neanderthalis* and *homo sapiens*—living together in a single moral community. (Modulo the point about not interbreeding, some anthropologists think that this has actually happened: Trinkaus and Shipman 1993.) Pace Carruthers, his argument does not show that there is no species-classification that bestows participation in the moral community on its members, and so that species-classifications are morally irrelevant. What his argument shows is that there may be *more than one* species-classification (or other kind-classification) that bestows participation in the primary moral constituency on its members. That is something that we should grant anyway given the possibility of other species whose members are characteristically persons; and it is not inconsistent with the claim that species-classifications are morally significant.<sup>xxvi</sup>

*d. Species and Darwinism.* A fourth objection to my argument is the claim that species don't exist: that as a matter of strict biology, there is no clear differentiation between humankind and "other species", only a differentiation between different parts of the same evolutionary family tree. Species therefore are populations, not Aristotelian (or Kripkean) natural kinds (Sober 1994). True, but irrelevant to my argument, which can work equally well with the population conception of species, or indeed with any conception at all that will allow "humanity" to count as the name of a kind of *some* kind. At least for creatures like us, whose temporal experience runs only over decades, hardly ever over centuries even, let alone the millions of years that speciation takes, it should not be a controversial thought that a population of creatures sharing a common genetic heritage, physiology, and ethology can sensibly be treated as a unitary grouping, as in fact common sense treats it. And that thought is all I am committed to meaning by "species".

So it looks like humanism can resist these four objections. And it looks like humanism can accept, indeed welcome, the thought that humans are not the only persons, the only members of the primary moral constituency—or at least, they might easily not be. What about the question on the other side: the question whether all human beings are persons? I turn to this question in the next and final section.

## **5. Are all human beings persons?**

Humanism as I have developed it here straightforwardly implies that all human beings are persons. Our treatment of any human being should be conditioned by the background of expectations, hopes, and aspirations that spell out what we know, from experience, humans in general can be. *Eudaimonia* in its broad outlines is the same for all human beings; and requires, as a general rule, that we must give all human beings the space to achieve *eudaimonia*—whether or not they predictably *will* achieve *eudaimonia*. To deny this space to any individual human being is to exclude that individual from the moral community of persons. And that is a serious injustice.<sup>xxvii</sup>

The very young are not excluded from the moral community of persons, just because they have not yet achieved many key forms of human *eudaimonia*; the mentally handicapped are not excluded from the moral community of persons, just because they cannot ever achieve many key forms of human *eudaimonia*. As I pointed out in section 1, *all* persons are incomplete and impure in their agency. No one ever



does everything she could with her own nature (partly, of course, because there are just too many things she could do). In one way or another, *all* of us are less than fully free to pursue happiness; everyone is wounded or damaged or limited or compulsive, or just plain ill, in some respect. That does not make it wrong to treat “normal” humans, whatever their age or IQ, with an interpretive charity, and with a kind of pity, that sees beyond their limitations, and leaves the door open, in each individual’s case, to what *any human* might become—even if we are well aware that *this person here* has no chance of becoming that.

[T]he kind of care that was needed to make us what we have in fact become, independent practical reasoners, had to be, if it was to be effective, unconditional care for the human being as such, whatever the outcome. And this is the kind of care that we in turn now owe or will owe. Of the brain-damaged, of those almost incapable of movement, of the autistic, of all such we have to say: This could have been us. (MacIntyre 1999: 100)

There is the possibility of deep moral concern for retarded people, in which they are seen as having, however incomprehensible we may find it, a human fate, as much as anyone else’s.... Someone may be very touched by the response of a severely retarded person to music; and there may be in that being touched an imaginative sense of shared humanity. (Diamond 1991: 55)

The terminally ill are not an exception to this rule either, just because they cannot any longer achieve many key forms of human *eudaimonia*. Perhaps it is true of them that they no longer have—to reuse Marquis’ phrase—“a future like ours”, and so will not achieve very much, or any, more *eudaimonia* before they die. Given that their futures have closed in on them in this radical way, you might call it “futile” to go on treating them with the full range of interpersonal hopes and expectations that you extend to any human, or to go on being as scrupulous about not killing them as you would be about anyone else. There again, you might also call it “respectful”; you might call it an expression of *pity*.

To pity someone in the sense I have in mind is to think, as MacIntyre puts it, “this could have been us”: this person too is part of human life and shares in human ideals, even if the idea of realising them in almost any way is quite hopeless for him/her. Such pity is a key part of our charity towards other people; of what is involved in recognising them as *people*, as fellow human beings.

(It is interestingly not something that follows in any very clear way from an Aristotelian virtue ethics where the focus is, eudaimonistically, on *maximal* forms of flourishing—on how much you as a being can develop, not on how many relatively undeveloped beings you can extend charity towards. Indeed it is a notorious fact about Aristotle that there is something rather ruthlessly bright-eye-and-gleaming-coat<sup>xxviii</sup> about his conception of *eudaimonia*, and that—not unconnectedly—he seems to have had little room in his ethics for the notion of pity.<sup>xxix</sup> Pity is, however, part of the sorts of virtue ethics that we may find in a Christian writer like Simone Weil. And that is the sort that I would defend, though I cannot defend it here.)

There are, as many people have emphasised, better and worse ways to die. And one of the worse ways to die is to have recognition of your humanity withdrawn

from you before it happens. Part of understanding why that is such a bad thing to have happen to you must, I think, depend on taking a different view from the criterialist's about the locus of the value of persons. For the criterialist, apparently, the value of persons consists in their instantiating the person-constituting properties. Aside from such instantiations, persons, at the most basic level, *have* no value (compare the familiar view in bioethics that being alive in itself is neither good nor bad); indeed there is a sense in which, aside from such instantiations, there *are* no persons. What such a view of the value of persons obscures from view is the notion of loving an individual in itself; it seems to be only the *properties* of an individual that a consistent criterialist can love—and then, only when they are duly manifested. But the idea that the object of love is always and necessarily some property-instantiation, common though it is, is a dogma. If property-instantiations can be objects of love (we may fairly ask), why not things in other metaphysical categories: individual human beings, for example?<sup>xxx</sup>

Alongside this dogma, perhaps there is another, related but slightly different, dogma at work in securing the foundations of criterialism. This is the dogma that whenever we answer the question “Why is X valuable?” by citing some valuable property V that X displays, what we must be really saying is always that it is (this instance of) V that is valuable, not X. But this too is a dogma, and unmasking it may be another way of setting ourselves free from the lures of criterialism. To cite V in explanation of the claim that X is valuable can, so to speak, be understood the other way around: you can take it that citing V is displaying a *symptom*, a *reminder*, of the value of X. When George Orwell, in a famous passage from his essay “A Hanging”, describes the growing fingernails and the refusal to walk through a puddle of the man who is about to be hanged, I take it that this latter thing is what he is doing.<sup>xxxi</sup> He is not saying that having growing fingernails is a value-giving property. Rather he is saying that *being human* is a value-giving property, the importance of which can be brought out for us by all sorts of exercises in the “assembling of reminders”—including such reminders as the growth of fingernails.

“But all the same, isn't there a profound difference between what we hope for *in general* for other people, and what we hope in the particular case of someone who is severely mentally handicapped?” There certainly is. There is always a distinction between the aspirations for other persons that are grounded in human *eudaimonia* in general, and the aspirations for specific other persons that are grounded in our particular knowledge of them as individuals. This distinction is just as applicable to the profoundly mentally handicapped as it is to anyone else. And it makes for the same commonalities, and the same differences, as with anyone else. If your friend is tone-deaf, there is no need to pretend that he will ever be Mozart; but that does not make it all right to take home all his CDs. If your friend loses both his legs in a skiing accident, you needn't spend the rest of his life pretending to him that he can still ski; but you miss something vital, too, if you bounce into his hospital ward straight after his double amputation with the words “So, you won't be wanting those ski-boots any more”. Many of the details of what counts as just and loving treatment of other persons are, certainly, set by the particular trajectories of their particular lives. But the most basic and elementary requirements of love and justice, is set by reference to the aspirations that arise from human *eudaimonia* in general. And this applies just as much to the mentally disabled as to anyone else.

In Nick Hornby's very funny novel about suicidal depression, *A Long Way Down*, Maureen, the mother of a severely mentally disabled child, says this:

This all began years ago, when I decided to decorate his bedroom. He was eight, and he still slept in a nursery—clowns on the curtains, bunny rabbits on the frieze round the wall, all the things I'd chosen when I was waiting for him and I didn't know what he was. And it was all peeling away... and I hadn't done anything about it because it made me think too much about all the things that weren't happening to him, all the ways he wasn't growing up. What was I going to replace the bunny rabbits with? He was eight, so perhaps trains and rocket ships and footballers were the right sort of thing for him—but of course, he didn't know what any of those things were... But there again, he didn't know what the rabbits were either, or the clowns. So what was I supposed to do? Everything was pretending, wasn't it? The only thing I could do that wasn't make-believe was paint the walls white, get a plain pair of curtains. That would be a way of [saying] that I knew he was a vegetable, a cabbage, and I wasn't trying to hide it. But then, where does it stop? Does that mean you can never buy him a T-shirt with a word on it, or a picture, because he'll never read, and he can't make sense of pictures? And who knows whether he even gets anything out of colours, or patterns? And it goes without saying that talking to him is ridiculous, and smiling at him, and kissing him on the head. Everything I do is pretending, so why not pretend properly?

In the end, I went for trains on the curtains, and your man from *Star Wars* on the lampshade. And soon after that, I started buying comics every now and then, just to see what a lad of his age might be reading and thinking about. And we started watching the Saturday morning television together, so I learned a little bit about pop singers he might like... thinking about these things helped me to see Matty, in a strange sort of way... I made up a son. (Nick Hornby, *A Long Way Down*, pp.118-9)

Of course there is—as Maureen herself recognises—something delusional, and something pathetic, in her treatment of Matty. For sure, her attitude to him is strained and exaggerated. But what we should not lose sight of is that it is a strained and exaggerated version of something morally indispensable: the kind of interpretive charity, the humanistic understanding and pity, that, I have argued here, we all need from everybody else all the time. In the words of an older writer: “Charity bears all things, believes all things, hopes all things, endures all things.”<sup>xxxiii</sup>

## Bibliography

- Bauby 1998: Jean-Dominique Bauby, *The Diving Bell and the Butterfly*. London: Fourth Estate, 1998.
- Carruthers 1992: Peter Carruthers, *The Animals Issue: Moral Theory in Practice*. Cambridge: Cambridge University Press.
- Chappell 1998: Timothy Chappell, *Understanding Human Goods*. Edinburgh: Edinburgh University Press.
- Chappell 2003: Timothy Chappell, "Absolutes and particulars", 95-117 in *Modern moral philosophy*, ed. A. O'Hear. Cambridge: Cambridge University Press.
- Davidson 1980: Donald Davidson, *Inquiries into truth and interpretation*. Oxford: Clarendon.
- Darwall 2006: Stephen Darwall, *The Second-Person Standpoint*. Cambridge, MA: Harvard University Press.
- Diamond 1991: Cora Diamond, „The importance of being human“, in David Cockburn, ed., *Human Beings*. Cambridge: Cambridge University Press.
- Engelhardt 1989: Tristram Engelhardt, "Ethical Issues in Aiding the Death of Young Children", in *Euthanasia: the moral issues*, edd. Robert M Baird and Stuart E Rosenbaum. Buffalo, N.Y. : Prometheus Books.
- Gaita 1991: Raimond Gaita, *Good and Evil: an absolute conception*. London: Routledge.
- Gaita 1998: Raimond Gaita, *A Common Humanity*. London: Routledge.
- Grau 2010: Christopher Grau, "Love and history", *Southern Journal of Philosophy* **48** (2010), 246-271.
- Grice 1957: H.P.Grice, "Meaning", *The Philosophical Review*, **66**: 377-388.
- Harris 1985: John Harris, *The Value of Life*. London: Routledge.
- Alison Hills, "Utilitarianism, contractualism and demandingness", *Philosophical Quarterly* **60** (April 2010), 225-242.
- Nick Hornby, *A Long Way Down*. London: Penguin, 2006.
- David Hume: *A Treatise of Human Nature* [1739].
- Kahane 2010: Guy Kahane, "Feeling pain for the very first time: the normative knowledge argument", *Philosophy and Phenomenological Research* **80.1** (2010), 20-49.
- Long and Sedley 1987: A.A. Long and D. Sedley, *The Hellenistic Philosophers*, Cambridge University Press, Cambridge, Vol.I, 67A
- Marquis 1997: Don Marquis, "An argument that abortion is wrong", in Hugh LaFollette, *Ethics in Practice*, First Edition (Oxford: Blackwell, 1997).
- MacIntyre 1999: Alasdair MacIntyre, *Dependent Rational Animals: why human beings need the virtues*. Chicago: Open Court, 1999.
- McMahan 2005: Jeff McMahan, "Our fellow creatures", *Journal of Ethics* **9** (2005), 353-380.
- Olson 2007: Eric Olson, Stanford Encyclopedia of Philosophy entry on "personal identity", <http://plato.stanford.edu/entries/identity-personal/>
- Parfit 1984: Derek Parfit, *Reasons and Persons*. Oxford: Oxford University Press.
- Patterson and Gordon 1993: "The Case for the Personhood of Gorillas", in Paola Cavalieri and Peter Singer (eds.), *The Great Ape Project* (New York: St. Martin's 1993), 58-77.
- Rachels 1990: James Rachels, *Created From Animals: The Moral Implications of Darwinism*. Oxford: Oxford University Press.
- Schapiro 1999: Tamar Schapiro, "What is a child?", *Ethics* **109.4**: 715-738.

- Singer 1983: Peter Singer, *The Expanding Circle*. Oxford: Oxford University Press.
- Singer 1991: Peter Singer, *A Companion to Ethics*. Oxford: Blackwell.
- Singer 1993: Peter Singer, *Practical Ethics*. Cambridge: Cambridge University Press.
- Singer 2002: Peter Singer, *Unsanctifying Human Life: essays on ethics*, ed. Helga Kuhse. Oxford: Blackwell.
- Sober 1994: Elliott Sober, *From a Biological Point of View*. Cambridge: Cambridge University Press.
- Strawson 1962: Peter Strawson, 'Freedom and Resentment', in his *Freedom and Resentment and other Essays*, London: Methuen, 1974.
- Teichman 1996: Jenny Teichman, *Social Ethics*. Oxford: Blackwell's.
- Tooley 1972: Michael Tooley, "Abortion and Infanticide", in Singer, ed., *Applied Ethics* (Oxford: OUP, 1987).
- Vlastos 1981: Gregory Vlastos, "The Individual as Object of Love in Plato", in his *Platonic Studies*, Princeton, NJ: Princeton University Press, 3–42, 2nd edn.
- Warren 1997: Mary Anne Warren, "On the Moral and Legal Status of Abortion", in H.Lafollette, ed., *Ethics in Practice* (Oxford: Blackwell, 1997).
- Williams 1985: Bernard Williams, *Ethics and the Limits of Philosophy*. London: Penguin.
- Williams 2006: Bernard Williams, *Philosophy as a humanistic discipline*. Cambridge: Cambridge University Press.
- Williams 2006a: Bernard Williams, "Hylomorphism", 218-230 in his *The Sense of The Past*. Cambridge: CUP.
- Ludwig Wittgenstein, *Philosophical Investigations*. Oxford: Blackwell, 1951.

## Notes

---

<sup>i</sup> Anscombe mistranslates *zur Seele* as “to a soul”; I have corrected this to “to *the* soul”.

<sup>ii</sup> Perhaps we could do ethics entirely or almost entirely without using the word or concept “person”: so numerous philosophers have at times suggested, including Jenny Teichman (1996), Bernard Williams (2006a), Christopher Coepe (2006), and myself (I have at least had this thought; I can’t now remember if it has ever seen the light of published day; something close to that thought is certainly at work in Chappell 1998, Chapter 4). However, I don’t pursue that idea here.

<sup>iii</sup> Criterialism, as I call it, is pretty much the same view as what McMahan 2005, following Rachels 1990, calls “moral individualism”. This is the view (Rachels 1990: 173) that “how an individual may be treated is determined, not by considering his group memberships, but by considering his own particular characteristics”.

<sup>iv</sup> Or something very close to criterialism. I need not commit myself yet on whether these authors are defending necessary-and-sufficient-conditions versions of a criterial view.

<sup>v</sup> Something like the criterial view pops up in many other places. For instance, it seems to be behind Olson 2007’s suggestion that “person” is a phase sortal: “Being a person may be only a temporary property of you, like being a philosopher”.

<sup>vi</sup> It might be suggested that (some instance of) Criterialism is a *bespoke* view of personhood, aimed only at defining what persons are for some specific purpose. Such a position is perfectly possible, of course, but I think it is very clearly *not* what is at issue in discussions like the four I cite here, all of which are aimed at completely general conclusions about “what persons are”: what they are in their essence, and for *any* possible purpose. (Thanks to an anonymous referee for discussion of this point.)

<sup>vii</sup> For more on personhood and language, see Patterson and Gordon 1993.

<sup>viii</sup> David Hume: “For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch *myself* at any time without a perception, and never can observe any thing but the perception. When my perceptions are removed for any time, as by sound sleep, so long am I insensible of *myself*, and may truly be said not to exist” (*Treatise* 1.6.3). Derek Parfit: “We could give a complete description of reality without claiming that persons exist” (Parfit 1984: 212). The Buddha: “There exists no individual, it is only a conventional name given to a set of elements” (quoted in Parfit 1984: 502).

<sup>ix</sup> Autism involves very various sorts of emotional dysfunction. It can involve *over*-emotionality and “inappropriate” emotional responses as well as a lack of emotional response. But that just underlines the absurdity of using “emotionality” as a necessary condition of personhood. For then autistics who are over-emotional will count as persons, autistics who are “cold” and “withdrawn” won’t.

<sup>x</sup> There is a surprising contrast between some recent philosophers’ willingness to grant moral standing to “potential persons” in the sense of metaphysically possible future human beings who admittedly do not exist right now, and their reluctance to do so to “potential persons” in the sense of young human beings who admittedly do exist right now.

<sup>xi</sup> “The wrongness of killing us is understood in terms of what killing does to us. Killing us imposes on us the misfortune of premature death. That misfortune underlies the wrongness... The misfortune of premature death consists in the loss to us of the future goods of consciousness... What makes my future valuable to me are those aspects of my future that I will (or would) value when I will (or would) experience them... What makes killing us wrong, in general, is that it deprives us of a future of value. Thus, killing someone is wrong, in general, when it deprives her of a future like ours.” (Marquis 1997: 95-96)

<sup>xii</sup> And it is nearly always *the* criterion, i.e. some *one* criterion, that we are talking about. (Warren’s list of six criteria is the only exception to this monism that I can think of.) The suspicion that the value of persons might be multiform not monistic, that there might be not just one simple and readily-articulable reason for valuing persons but *all sorts* of reasons, some of them far from simple or easy to put into words, is not at all a characteristic suspicion for criterialists.

<sup>xiii</sup> About the argument that I develop from this point on, about “how we treat” people, it is natural to ask: are these descriptive or normative claims? The answer to that, as quite often with Wittgensteinians like (I suppose) me, is: a bit of both. What is on offer is a phenomenology of (part of) the moral life. My claim is that the most basic and fundamental truths about this phenomenology makes certain options in thinking about ethics that might seem attractive and viable, such as criterialism, actually incoherent, because of its mismatch with what is basic and fundamental. Because of how things *are*, there are some things we *shouldn’t* think or do: both a normative and a descriptive thesis.

---

Another worry that might occur to some here: am I attacking criterialism because it is a *theory* of personhood? If so, isn't it hypocritical of me to go on to develop what is really just another theory of personhood? The answers are No and No. I'm not attacking criterialism because it's a *theory*, but because it's a *rotten* theory. A good theory (or if you dislike that word, a good philosophical account) of personhood will make good sense of our intuitions; this criterialism signally fails to do.

(My thanks to anonymous referees for pressing these questions.)

<sup>xiv</sup> The charming comparison is Singer's: Singer 1993: 89-90.

<sup>xv</sup> But when exactly? As part of the child's fourth birthday celebrations? On going to primary school? Would a Singerian society have a ritual for an induction into personhood?

<sup>xvi</sup> At least, the term "objective" is unfortunate if it implies, as I suspect it does in passages like the one from Strawson that I cite in the next Note, that there is something *non-objective* about participant reactive attitudes.

<sup>xvii</sup> "We look with an objective eye on the compulsive behaviour of the neurotic or the tiresome behaviour of a very young child, thinking in terms of treatment or training. But we sometimes look with something like the same eye on the behaviour of the normal and the mature. We have this resource [the objective attitude] and can sometimes use it. Being human we cannot, in the normal case, do this for very long, or altogether. But what is above all interesting is the tension there is, in us, between the participant attitude and the objective attitude. One is tempted to say: between our humanity and our intelligence. But to say this would be to distort both notions." (Strawson 1962: 66-67, with omissions)

<sup>xviii</sup> "One speaks and expresses emotions to and not simply at one's cat. Nonetheless, although we of course discipline them, we don't press claims against or hold our pets accountable in the same way we do with one another... Granted, when I look into my cat's eyes, I can't get over the feeling that he is looking also into mine in some personal way or shake the hope that he is seeing me in some way other than 'the guy who feeds me'. But I find it utterly impossible to sustain the thought that he can imaginatively enter into my point of view or acknowledge me as a being with an independent perspective." (Darwall 2006: 43)

Pets are an interesting case of something like a "pretend" interpersonal relationship, if "pretend" is the right thing to call it. We know that, unlike infants, our cats will *never* become responsive in the way that persons do. That doesn't stop us (or at least the dedicated pet-owners among us) from extending something interestingly like the constitutive prolepsis even to cats, hamsters, or stick-insects. Children, of course, do something similar with their dolls or other toys—and here "pretend" clearly is the right word. For a recent fashion to which similar comments apply see <http://www.nytimes.com/2009/07/26/magazine/26FOB-2DLove-t.html> . (Thanks to an anonymous referee for drawing this to my attention.)

<sup>xix</sup> "Intrinsic" in the sense of "essential", I mean; not in the sense of "non-relational", obviously.

<sup>xx</sup> Gaita 1991: 117: "An animal can suffer, but it cannot curse the day it was born. An animal can be afraid, but it cannot be ashamed of its fear and despite itself. An animal can be happy, but it cannot be joyous. An animal cannot give of its substance to certain pursuits and be admonished for doing so. One could go on almost indefinitely. The problems of life's meaning cannot arise for an animal. Only a being for whom life can be problematic can have a spiritual life, and therefore have a soul."

<sup>xxi</sup> For more examples applying this argument see Hills 2010: 235-7.

<sup>xxii</sup> So I say. In the case of trees, of course, there was at least one stern Victorian moralist who thought otherwise, namely W.E.Gladstone, whose hobby was chopping down trees.

<sup>xxiii</sup> For this speculation deployed for a different purpose, cp. Kahane 2010.

<sup>xxiv</sup> Cp. Gerald the gorilla in the well-known *Not the Nine O'Clock News* sketch:

<http://www.youtube.com/watch?v=MpbMm0433I>

<sup>xxv</sup> "Suppose that the experiments attempting to teach language to chimpanzees had been successful beyond their originators' wildest dreams. The apes in question gained a complete mastery of English within a few years, were able to attend school and later university, and made many close friendships with human beings... in these imaginary circumstances it would plainly be absurd to claim that the apes lacked moral standing, or had a moral importance that was lower than our own." (Carruthers 1992, Ch.3)

<sup>xxvi</sup> Carruthers' own view about animal rights is much closer to speciesism than to animal liberationism. He holds that there is an indirect argument, arising from contractualism, for granting all humans and no animals moral standing (for a brief exposition see his online paper "The animals issue" at <http://www.philosophy.umd.edu/Faculty/pcarruthers/The%20Animals%20Issue.pdf>). Since I think it is obviously true that (non-human) animals do have moral standing in their own right, based roughly on the wrongness of causing unnecessary suffering, damage, and destruction, it seems I am closer to animal liberationism than Carruthers is.

---

<sup>xxvii</sup> Jeff McMahan (2005: 366) writes that his “moral individualism” gives us reason to deny that it is a grave misfortune for a human individual who is disabled to have limited cognitive capacities. After all, it is not a misfortune for an individual animal to have capacities of roughly the same level; and according to moral individualism, it is only the individual creatures that we should be considering. So either both the human and the animal are unfortunate, or neither is.

This seems a very counter-intuitive consequence of “moral individualism”. Surely it matters how the creatures got into their states: at the very least, there must be a difference in fortune between a human individual who was born at a low cognitive level, and one who is at that level because of some horrendous accident? So far as I can see, McMahan must deny this. For presumably he thinks that my past stands to me as a relational characteristic, not an intrinsic one—and the whole point of “moral individualism” is that only “intrinsic characteristics” count morally.

<sup>xxviii</sup> Bernard Williams’ delightfully mordant phrase: Williams 1985: 46.

<sup>xxix</sup> The best presentation I know of the evidence for Aristotle’s blindness on the subject of pity is in Gaita 1991.

<sup>xxx</sup> For more on this point see Chappell 2003, Grau 2010, Vlastos 1981.

<sup>xxxi</sup> “It is curious, but till that moment I had never realized what it means to destroy a healthy, conscious man. When I saw the prisoner step aside to avoid the puddle, I saw the mystery, the unspeakable wrongness, of cutting a life short when it is in full tide. This man was not dying, he was alive just as we were alive. All the organs of his body were working—bowels digesting food, skin renewing itself, nails growing, tissues forming—all toiling away in solemn foolery. His nails would still be growing when he stood on the drop, when he was falling through the air with a tenth of a second to live. His eyes saw the yellow gravel and the grey walls, and his brain still remembered, foresaw, reasoned—reasoned even about puddles. He and we were a party of men walking together, seeing, hearing, feeling, understanding the same world; and in two minutes, with a sudden snap, one of us would be gone—one mind less, one world less.” (George Orwell, “A Hanging”, *Adelphi Magazine* August 1931)

<sup>xxxii</sup> My thanks for their comments and encouragement to a large number of people, including Julia Annas, Alex Barber, Chris Belshaw, David Benatar, Peter Cave, Stephen Clark, John Cottingham, Christopher Coope, Christopher Grau, John Haldane, Rosalind Hursthouse, Tim Mulgan, David Oderberg, Tim O’Hagan, Charles Taliaferro, Chris Tollefsen, and Garrath Williams; an audience in Leeds at the Society for Applied Philosophy in 2009; and three anonymous referees for this journal.



## **On the very idea of criteria for personhood**

### *Abstract*

I examine the familiar criterial view of personhood, according to which the possession of personal properties such as self-consciousness, emotionality, sentience, and so forth is necessary and sufficient for the status of a person. I argue that this view confuses criteria for personhood with parts of an ideal of personhood. In normal cases, we have already identified a creature as a person before we start looking for it to manifest the personal properties, indeed this pre-identification is part of what makes it possible for us to see and interpret the creature as a person in the first place. And that pre-identification typically runs on biological lines. Except in some interesting special or science-fiction cases, some of which I discuss, it is human animals that we identify as persons.

## On the very idea of criteria for personhood

Meine Einstellung zu ihm ist eine Einstellung zur Seele. Ich habe nicht die *Meinung*, dass er eine Seele hat.

My attitude towards him is an attitude towards the<sup>i</sup> soul. I am not of the *opinion* that he has a soul. (Ludwig Wittgenstein, *Philosophical Investigations*, II, p.152, tr. G.E.M.Anscombe)

Love is conditioned by its object, but love also yields its object. (Gaita 1991: 122)

### 1. Persons: humanism, speciesism, criterialism

Any ethical outlook much like ours will take as central some *primary moral constituency* (PMC): some class of creatures who all alike, and all equally, share in the highest level of moral rights and privileges. Most philosophical ethicists use “person” to mean at least “member of the primary moral constituency” (whatever else they may also mean by “person”). In this paper, that is how I shall use the word too.<sup>ii</sup>

So what makes any creature a person, a member of the primary moral constituency? One view is what I shall call humanism:

**Humanism:** Being human is sufficient for membership of the PMC.

Humanism faces two very common objections. The first is that humanism is arbitrary discrimination, like racism, sexism, ageism; humanism, in Richard Ryder’s famous phrase, is *speciesism*. The second objection, which is really just the same idea taken a little deeper, says that humanism bases its demarcation of persons on a biological property which is morally insignificant. Membership of this or that species—just like membership of this or that race, gender, or age-group—is, we are told, a morally insignificant detail. It is not the kind of thing that we should expect any important moral distinction to be based on.

In theory these objections could be put by anyone who rejected humanism. In practice they are usually put by proponents of another theory of personhood, which I shall call criterialism:

**Criterialism:** Actual possession of the criterial properties is necessary and sufficient for membership of the PMC.<sup>iii</sup>

Here are four very well-known enunciations of criterialism<sup>iv</sup>:

I propose to use “person”, in the sense of a rational and self-conscious being, to capture those elements of the popular sense of “human being” that are not covered by “member of the species *Homo sapiens*”. (Singer 1993: 87)

Persons are beings capable of valuing their own lives. (Harris 1985: 16-17)

An organism possesses a serious right to life only if it possesses the concept of a self as a continuing subject of experiences and other mental states, and believes that it is itself such an entity. (Tooley 1972: 82)

[The six key markers of personhood are] (1) sentience... (2) emotionality... (3) reason... (4) the capacity to communicate... (5) self-awareness... (6) moral agency. (Warren 1997: 83-84)<sup>v</sup>

As here represented, the criterial view is an all-purpose<sup>vi</sup> view about what personhood is, and why it matters, which is undeniably simple (as I shall argue, much too simple). Despite this simplicity, it is obvious, when criterialism is stated as baldly as it is in these four statements, that it has radical implications. Most noticeably, it implies a sharp contraction of the circle of intuitive moral concern. If rationality is a necessary condition of counting as a person, a member of the primary moral constituency, then all children under a certain age will fail to count as persons. (This bullet is explicitly bitten by Tristan Engelhardt: “If being a person is to be [*sic*] a responsible agent, a bearer of rights and duties, children are not persons in a strict sense” (Engelhardt 1989: 120).) If self-consciousness is a necessary condition, then many mentally handicapped humans will fail to count as persons. If the capacity to communicate is a necessary condition, then Jean-Dominique Bauby, of *The Diving Bell and the Butterfly* fame, would have failed to count as a person after his catastrophic stroke—had he not worked out how to signal with his eyelid.<sup>vii</sup> (Can whether Bauby is still a person after his stroke really depend on contingencies about whether he can still communicate?) If persons necessarily have “the concept of a self as a continuing subject”, and believe that they are such selves, then David Hume, Derek Parfit, and most Buddhists are not persons.<sup>viii</sup> If you cannot be a person unless you are capable of valuing your own life, then you cannot be both a person and a committed nihilist about value (i.e. someone who finds himself compelled to believe that no such thing as value exists). If persons have to possess “emotionality”, then Mr Spock in *Star Trek* is not a person, and neither, perhaps, are some extreme autistics.<sup>ix</sup> And so on. Criterialists can make personhood sound rather exclusive; it can seem as hard to qualify for personhood as it is to make membership of the Country Club.

Of course, these problems are consequences of taking these “markers of personhood” as indicators of necessary conditions for personhood. A less extreme version of the criterial view might take each of the markers on some list, or some or other combination of the markers, as a sufficient condition of personhood: any creature that displays just one of these properties, or some large-enough variety of them, will count as a person. We might then want to argue about what counts as displaying a property, or possessing a capacity. Must I actually evince emotions to pass Warren’s emotionality test? How often, and how convincingly? Do I possess the capacity to communicate only when I have learned a language? Or just when I have learned to sign, or to get others to read my thoughts and feelings? Or do I have the capacity to communicate all along, just in virtue of being a member of a species that communicates, linguistically and in other ways?

Here we touch on familiar and long-running debates about potentiality versus actuality in our assessments of personhood. It is commonplace for criterialists of Singer’s sort to point out that potential prime ministers obviously do not have the rights of actual prime ministers, so that “potential persons”<sup>x</sup>, or those with the

potential for what Don Marquis famously called “a future like ours”<sup>xi</sup>, can hardly be treated as the same moral category as actual persons, or “us”. It is equally commonplace for the opponents of Singer-style criterialism to object that there must be something wrong with an emphasis on *actual* properties that makes it a live question whether one is a person while asleep or under general anaesthetic.

Clearly there are problems about at least some of the suggestions listed above about what the criteria of personhood might actually be. But not all those suggestions are implausible, and it would be easy enough, in principle, to weed out or refine the implausible ones. True, it is striking how little criterialists feel they need to do this; in practice, criterialists often seem perfectly happy to stick with what, intuitively, seem implausibly over-demanding criteria for personhood.

However, the plausibility or otherwise of the various suggested criteria is not my main concern here. Obviously enough, other criteria could be devised which did not have these consequences. The key difficulties for criterialism that I want to emphasise in this paper are three.

The first, which I can state very quickly, is about the argumentative method underlying criterialism. To put it at its simplest, it *looks* as if criterialists typically adopt their preferred criterion of moral significance because they think it explains moral intuitions which we have anyway. The criterion is adopted, in other words, because it doesn't have counter-intuitive consequences.<sup>xii</sup> But then, the same criterion is deployed, by the very same authors—Peter Singer and John Harris, for example—precisely to *produce* counter-intuitive consequences, such as the moral acceptability of killing small babies or the mentally disabled. There is, I think, a serious question as to what can possibly be going on here. It is not entirely obvious why building a moral theory out of one intuition or set of intuitions that we certainly have, and then using this theory to attack other intuitions that we also certainly have, is anything more than arbitrary selectiveness. Perhaps this is a question that all system-building moral theory faces.

The second difficulty for criterialism, which I shall also mention only very briefly, is that criterialism can seem to imply a quite implausible picture of moral reasoning. Are we really to suppose that what we do when we come across any being we haven't met before is *first* see whether it satisfies our criterion for personhood, and *then*, if it does, start treating it as a person? One is reminded of that well-known caricature the Cartesian Detective: we start with the hypothesis—the *opinion*, as Wittgenstein says in my epigraph—that the new being is a person, and feel justified in acting in line with that hypothesis, just insofar as the behavioural evidence confirms it. If criterialism is to be at all plausible, it cannot in this way be a general view about our typical real-time interaction with others. Such a picture might apply to some very exceptional cases, some of which we will discuss in section 4b. But in general, criterialism will have to be a view about the *criterion of rightness* for taking other creatures to be persons, not about *our decision procedure* for assuming that they are persons. For quite obviously, our normal interaction with others typically involves no such Cartesian-Detective rigmarole, and it takes a good deal of philosophical callousing of the soul not to find something absurd, and morally repugnant too, about the idea that it even could. But if in general criterialism only states the criterion of rightness, and not our decision procedure, for taking other creatures to be persons,

then it still faces a further objection, my third—which it is the main purpose of this paper to develop.

This third objection to criterialism, on which I shall spend the rest of this paper, is connected to this second point. My claim is that the sort of properties that criterialists home in on are not criteria of personhood at all. Rather, they are dimensions of interpretation of beings that we *already* take to be persons.

For my own part, I have always been aware of at least a vague sense of discomfort and unease here, the moment we start talking about these (or any other) properties as if they were criteria for personhood of the sort the criterialists have in mind. There is a hard-to-articulate feeling that, despite the considerable first-blush plausibility of criterialism, it leaves out something crucial; a feeling of the kind that is naturally expressed in ordinary conversation by a trailing-off “Yes, *but...*”. I suspect that my unease is not unrelated to the unease that Raimond Gaita is bringing out when he writes (Gaita 1991: 115) that it is

naïve to think that the kinds of properties which interest philosophers when they ask what is a person, and which we often share to some degree with higher animals, play the kind of role in our treatment of one another which is assumed by those who argue that differences in treatment must be justified by relevantly different properties of those kinds. That philosophical perspective from which we are encouraged to reassess our sense of how we might justifiably treat animals distorts and indeed cheapens our understanding of human life.

Gaita 1991 is a marvellous exploration of one philosopher’s attempt to make articulate that trailing-off “Yes, *but...*”. This essay offers another philosopher’s. My attempt is not necessarily the same as Gaita’s, but not necessarily in competition with it either.

And it is perhaps worth saying before I begin that what I am trying to do here, what Gaita was trying to do, is *hard*: hard to articulate, hard to state clearly, hard to see in the round, hard to apply. Again and again criterialism wins out in philosophers’ discussions of personhood because, conversely, it is dead easy; it is such a simple and straightforward view of personhood. There are plenty of spurious analogies between science and ethics around, but at any rate we can trust this one: no more in the case of the nature of the person than in the case of particle physics does the fact that a view is *simple* in any way improve its chances of being *true*.

## **2. The proleptic view of personhood**

Contrary to what criterialism seems to suggest, we do not look for sentience or rationality or self-awareness in a creature as a test to decide whether or not that creature counts as a person. It’s the other way round. Having once decided, on other grounds, that a creature *is* a person, we know that this makes it the kind of creature which is likely to display sentience, rationality, self-awareness and the rest of the personal properties. Hence we look for displays of these properties from the creature. That is to say, we treat it as a person *in advance* of any such displays.<sup>xiii</sup>

The biggest and clearest example of this is an activity that is absolutely central to human life: parenting. Parents are, of course, aware of the differences in rationality, linguistic capacity, self-representing ability, and so forth between young human beings at various ages and developmental stages. Nonetheless, a parent's attitude towards her child is always, basically, what Wittgenstein famously calls "an attitude towards a soul". She does not do what criterialism (especially if understood in decision-procedure terms) might seem to imply she should do: start out by treating her child as an inanimate object, like a sofa or a fridge or a rubber-plant, and grudgingly consent to adjust her attitude to it, one little step at a time, only as and when it proves itself more than inanimate by passing a succession of behavioural tests for at least having interests like a snail's perhaps,<sup>xiv</sup> and then later on for the interests of some more advanced kind of animal such as a puppy, and eventually for full criterialist personhood.<sup>xv</sup> If parents *did* treat their children in this almost behaviouristic fashion, the parents would be callous monsters, and the children would be basket-cases. In a world where parents generally accepted this sort of criterial view, and applied it directly in their parenting practice, even the best-adjusted of us would be a wolf-child.

Fortunately, however, parents are not criterialists, either in their decision procedure or in the criterion of rightness for ascriptions of personhood that they actually employ. Rather, a parent treats her child from the very beginning—and from before it is literally and actually true—as a creature that can reason, respond, reflect, feel, laugh, think about itself as a person, think about others as persons too, and do everything else that persons characteristically do. From the beginning her attitudes towards the child are not only "objective" (to use Strawson 1962: 67's slightly unfortunate term<sup>xvi</sup>): they also include what he famously calls "participant reactive attitudes", of just the same sort as she adopts towards anyone else.

Of course, as Strawson implies<sup>xvii</sup>, these participant reactive attitudes' scope may be more restricted in the case of a very small baby. The point is that they are there; and that like the baby—often indeed in advance of the baby—they *grow*. As Alasdair MacIntyre notes in a recent discussion of the development of practical reasoning in the individual person (1999: 90), ethically and psychologically adequate parenthood centrally involves "a systematic *refusal* to treat the child in a way that is proportionate to its qualities and aptitudes" (that is, its actual and non-idealised qualities and aptitudes, the kind of qualities and aptitudes that interest the criterialist).

The parent who says to her three day old baby, "What do you think? Should we give you some more milk now?", or to her three year old toddler "How kind of you to share your toys with your sister when she's sad!", is not deluded about a neonate's capacity for interpersonal practical reasoning, or a toddler's capacity for calculatedly empathetic altruism. But she is not making a sentimental joke either. She is treating the baby or the toddler proleptically, in the light of the ideal of personhood. By years of treating her children as creatures who "have the personal properties"—in the sense that interests the criterialist—she makes it true that they *are* creatures who have the personal properties in just that sense.

The child staggers across and plonks a book down on his sibling's lap; the parent's reaction is "How kind of you to let her share your book!" Is that, in fact, what the toddler was doing? The right answer to this question can be: "Yes it is—once the

parent has given this reaction.” The parent’s reaction is an interpretation of the child’s deed. Before the reaction, perhaps, there was no fact of the matter about what the little boy was doing: the child did not know himself what his action was. For all he knew about it, his action might as well have been simply a random sequence of bodily movements. But now that his mother offers her interpretation, the child, on his mother’s authority, learns to see himself a certain way. Because his mother frames his act as one with a certain meaning (as the giving of a gift), and because the child sees his mother as seeing his act this way, and because his mother sees the child as seeing her as seeing his act this way... (Grice 1957), the act *comes to have* that meaning.

Paul Grice’s famous work on meaning and intention is even more relevant to another clear example of this sort of prolepsis, namely language-learning. A baby begins by babbling—that is, by producing all the sounds the human mouthparts can make. Some of these sounds get a response because they are sounds that occur in the parents’ language, while others don’t because they are not. (The inability of many English adults to pronounce the Scots *ch* in *loch* or the Welsh *ll* in *Llanfair* is a learnt inability, not mirrored in English babies, who can and do make both sounds until conditioned not to by English adults.) The baby “homes in” on the sounds, and then on the patterns or combinations of sounds, that get a response. Then it learns to correlate particular patterns of sounds with particular contexts by the same homing-in process. What the parents say is “She’s learned to say ‘tiger’” (my own first child’s first word). What has happened is that the parents, by treating their daughter *as if she has* the word for, and then the concept of, a tiger, have made it true that she *has* that word. (And, a little later perhaps, that concept too; though I am not suggesting that any one word or concept can be learned on its own: in Wittgenstein’s familiar image, “light dawns gradually over the whole” (*On Certainty* 141).)

Stephen Darwall too talks of prolepsis when describing a third connected process, namely the process whereby we come to be, and to hold others, morally responsible (Darwall 2006: 87-88):

There seem to be many cases where we wish to hold others accountable though we seem to have very good evidence that they are not free to act on moral reasons in the way our practices of holding someone fully responsible seem to presuppose... In some instances, for example with children, we seem simultaneously to move on two tracks in the process of inducting them into full second-person responsibility, sometimes treating them proleptically as though they were apt for second-personal address as a way of developing moral competence while nonetheless realising... that [at their developmental stage] this is an illusion.

Or an illusion, at any rate, if you agree with the criterialist about what it takes to be “apt for second-personal address”, and/or about what it takes to “have the personal properties”. My point, with which Darwall might or might not agree—I’m not sure—is that what it takes does not have to be already-realised and –actualised capacities of the sorts that interest criterialists. (The mother who asks her neonate child “Shall I give you some milk?” is certainly engaging in second-person address, *despite* the manifold incapacities, perfectly well-known to her, of neonates.)

There is at least this much truth in Dennett 1987's well-known idea of the "intentional stance": persons constitute each other as persons, both as agents and as patients, by *treating* each other as persons, as agents and patients. Something like the Davidsonian "principle of charity" (Davidson 1980) is at work in our mutual interpretations. By charitably, and proleptically, interpreting the other as a person, I *make* him a person.

The looseness and ambiguity of the term "person" is evident when we say things like this. In ordinary language, "person" can certainly, at one end of its semantic range, mean something like what Singer and other criterialists want it to mean, a "rational and self-conscious being" (or whatever). Yet at the other end of its semantic range, the ordinary-language term can also be a synonym for "human animal". It seems to be part of the *point* of our person-concept to allow those young humans who are persons only in the "human animal" sense to enter into a continuum of meaning towards the far end of which they can become persons in much stronger senses.

Many things in our shared life can be seen in this sort of way as aspirations towards an ideal which, like any ideal, is never fully realised. The basic sort of idealisation, the sort that grounds attributions of personhood, is not the only sort. We ascribe the rights to freedom of assembly and freedom of speech to every human being, even though we know that small babies can't speak and don't assemble (not without help), and that acutely asocial loners may well never (or at least never again) actually exercise either right. The behaviour of parents, besides (as described) involving the person-constituting prolepsis in a particularly clear form, is also full of other similarly-patterned proleptic idealisations that might, on a criterialist's view, be regarded as quite unrealistic. Parents give their children the opportunities to play the cello, to learn Spanish, to use their monthly allowances sensibly, to be polite to Great Aunt Maud (...), even though the parents may be fully aware that every one of these opportunities will in all probability be spurned. In minimally decent jurisdictions the treatment even of hardened criminals displays the same pattern of basic *openness* to the convict, however much of an unrealistic idealisation this openness may seem to involve: the jurisdiction goes on offering the criminals opportunities to reform, even though it is as good as certain that these opportunities will not be taken up. (I think that something like this refusal to "close the door" finally and irrevocably on the criminal is also, in connection perhaps with some thoughts about the state's obligation to be fundamentally benevolent to the citizens for whose sake alone it after all exists, the ground of a good reason to reject capital punishment. But that's another story.)

In general, the idealisation that accords X the status due to a  $\phi$ er does not have to involve any kind of expectation that X will ever actually  $\phi$ , or even can  $\phi$ . What it *does* have to involve is the understanding that X is of a kind that characteristically includes  $\phi$ ers. The human kind includes individuals who characteristically (at the right point in the human life-cycle) will freely assemble and speak. That is why it is not ridiculous to give the rights to free assembly, speech, and political action, e.g. becoming a senator, to any and every human, including asocial misanthropes, those (as yet, or as of recently) physically incapable of speech, and those who choose never to utter a word. Contrast the equine kind, which does not characteristically include any such individuals. That is why it *is* ridiculous to give these rights (or the status of a senator) to any horse.



Considered as a kind, children characteristically include individuals who learn the cello or Spanish or wise use of money or good manners to venerable relatives (well... eventually). That is why the parents who try to teach these lessons to their children are not acting absurdly even if they know in advance that *this particular* child is not going to respond—whereas parents who tried to teach these lessons to the family goldfish or the rubber plant *would* be acting absurdly. Similarly, there is nothing absurd about offering the old lag a place on the rehabilitation programme, even if it seems certain that *this* old lag will always refuse or abuse the offer: and the reason why not has to do with the general fact that some old lags do sometimes reform.

*A fortiori*, there is nothing absurd about giving the rights of free speech, assembly, private choice and the rest of it to women, or to other races, or to homosexuals or transgender people, or to the disabled, or to gingers, or to the various other sorts of historically marginalised groups. People in these categories are not another kind, as horses and hamsters are. Their well-being is *human* well-being, just as white male heterosexuals' is, and just as horses' and hamsters' well-being is not. That, in a nutshell, is why speciesism is *not* like racism or sexism.

Clearly, then, this is not just a point about the induction of babies and very small children into the social life of humanity at large. All humans, even the healthy intelligent mature well-adjusted independent affluent adults that typical criterialism is so clearly focused on—the Country Club again—are always only incompletely and impurely agents. Incompletely, because we are never all agents can be. However well things go, we never do or even can reach the full potential of human well-being, which would include a fully worked-out articulation of our own agency, a full coordination and marshalling of our own forces and powers for action in the world: towards this we are always only, at best, on the way. And impurely, because each of us always *can* be interpreted, by a sufficiently determined sceptic, as not really an agent at all, but a victim of the instinctive or other determining external forces that constantly threaten anyone's agency. To extend “an attitude towards a soul” towards *anyone* is always, and not just with children, an idealisation. As Tamar Schapiro notes within a Kantian framework (Schapiro 1999: 723):

[T]here is a sense in which no one, regardless of age or maturity, is able to achieve autonomy on Kant's view. This is because the notion of autonomy in Kant is an ideal concept which outstrips all possible realisations in experience. Strictly speaking, every instance of human willing is necessarily an imperfect realisation of transcendental freedom, and every virtuous character necessarily falls short of perfect virtue. And yet the applicability of the moral law depends on our mapping these ideal concepts onto ourselves and one another for the purposes of guiding action. So we are to regard the social world as a community of autonomous agents despite the fact that perfect realisations of autonomy are nowhere to be found.

The criterialists are right to insist that properties like self-awareness, emotionality, rationality, and so forth are crucial parts of our concept of a person. It does not follow that these properties can be used as the criterialists want to use them: as *tests* for personhood. Even at the cool-hour level of the criterion of rightness—never mind the

real-time level of deliberative procedure—to treat someone as a person is not to put a tick in the box by her name, to show that she has passed some inspection or met some standard, of rationality or self-awareness or emotionality or whatever. Indeed, it seems no less arbitrary discrimination to say “Sorry, you’re not rational or self-aware or linguistically capable or emotionally responsive enough to count as a person” than to say “Sorry, you’re not white enough to count as a person”. Behavioural properties like rationality, self-awareness, emotionality are not tests for, but parts of the ideal of, personhood. To treat someone as a person is to engage with him as the kind of creature to which that ideal applies. So to treat him is not, at the deepest level, a response to his *behaviour* at all, but to his nature. To see some creature as a person is to take an attitude to that creature which, *before* any behavioural evidence comes in, is already different from our attitudes to creatures that (we think) aren’t persons. This is the second-person standpoint (to use the title of Darwall’s outstanding book): the attitude we take to persons, and do not take to sofas, fridges, or rubber plants.

The fact that it takes charity, in more than one sense, to constitute persons imposes on persons a crucial sort of vulnerability and dependence: “The personal reactive attitudes rest on, and reflect, an expectation of, and demand for, the manifestation of a certain degree of good will or regard on the part of other human beings towards ourselves” (Strawson 1962: 70). All human persons are, in Alasdair MacIntyre’s splendidly apt phrase, “dependent rational animals”, as MacIntyre himself emphasises (MacIntyre 1999: 81-85):

In most moral philosophy the starting point is one that already presupposes the existence of mature independent practical reasoners whose social relationships are the relationships of the adult world. Childhood, if noticed at all, is a topic that receives only brief and incidental attention... To become an effective independent practical reasoner is an achievement, but it is always one to which others have made an essential contribution. The earliest of these relate directly to our animal existence, to what we share in our development with members of other intelligent species... [But what] we [also] need from others, if we are not only to exercise our initial animal capacities, but also to develop the capacities of independent practical reasoners, are those relationships necessary for fostering the ability to evaluate, modify, or reject our own practical judgements... Acknowledgement of dependence is the key to independence.

We are vulnerable to each other, and dependent on each other, because it is remarkably easy to attack persons by withdrawing the second-personal attitude from them. Part of the reason why it is so easy to stop seeing others as persons is, as I said before, because the interpersonal attitude always involves a degree of idealisation. There is always *some* truth in sceptical or cynical views that reject that idealisation. (For example, the human person *is* a physical object—even if it is not “merely” a physical object.)

We fail to see others as persons in a small way when we explain away what others do, or refuse to listen to them or to take their wishes and ambitions seriously. (Parents are never perfect in charity; children’s protests against such treatment are a recurring theme of family life.) We do it in bigger ways when we “use” others, or (in Kant’s phrase) fail to treat them as “ends in themselves”. At the limit, we can withdraw the interpersonal attitude from others completely, by pretending that they

simply aren't there, as British and Irish settlers did to the aboriginal Australians under their notorious legal doctrine of *terra nullius* (see Gaita 1998). Or we can do it by treating them merely as physical objects, as often happens in pornography, violence, and murder. Or, lastly, we can do it by declaring them "non-persons", as the Nazis did to the Jews—and as criterialists routinely do to the very old, the very young, and the very disabled.

The personal properties, then, set the content of the ideal of personhood. They do not set the boundaries of the class of persons (the PMC, as I've also called it). What does set the boundaries? That is my question in section 3. My answer will be the humanist answer already advertised in section 1: roughly, the boundaries of the class of persons are the boundaries of the human species.

### 3. The truth in speciesism

Rubber plants, fridges, sofas, human babies: we come to the last of these four sorts of things with a very different attitude from our attitudes to the other three sorts. And my point is that we *come to them* with different attitudes. Our attitudes to them are not responses to evidence that we gather *after* we have "come to them". *The moment I meet* a rubber plant or a fridge, I am perfectly sure, unless my sanity is in question, that it will not attempt to talk to me or tickle me—as, say, a baby might.

What makes me so sure of that? Well, simply the sort of thing a rubber plant or a fridge is, and the kinds of behaviour that it is natural to expect from that sort of thing. Rubber plants and fridges are not the kind of thing that tries to talk to you or tickle you. Babies are. And if you ask "What kind of thing is that?", the obvious answer is: "Young human beings". If some creature is a human being, it is completely natural to expect it to be the kind of thing that might (at least if it's a family member) try to talk to you, or tickle you, or otherwise demonstrate its (budding) mastery of the kinds of mental capacity that criterialists take to be criteria. (Talking to someone and tickling someone both require interesting interpersonal capabilities. To do either, you have to want things to be a certain way for someone else—roughly, you have to want to get them to think a thought, or to feel a tickle. These are wants with quite sophisticated contents.)

This sort of expectation is based squarely on the nature of the creature in question. And that means, on its nature *as a member of the human species*. Our understanding of which creatures are persons, and so may be expected to exhibit the personal properties, is not formed on an individual basis but on the basis of generalisation from experience of humans in general. I have already mentioned Wittgenstein's remark that "my attitude towards him is an attitude towards the soul"; on the same page of *Philosophical Investigations*, Wittgenstein also says that "the human body is the best picture of the human soul" (*der menschliche Körper ist das beste Bild der menschlichen Seele*). And perhaps what he means is this: that when we see that some creature is a human, just seeing that is enough to license us to adopt the interpersonal attitude to that creature.

We know in advance how we can expect any particular rubber plant or fridge to behave, because we can draw on a bank of past experience (our own and others') of

rubber plants and fridges. In this sense, the fact that a thing is a rubber plant or a fridge sets the scope of our expectations about it. In just the same way, we know what kinds of capacities and properties humans in general have. It is that background knowledge that we draw on, whenever we apply the interpersonal attitude to any individual human being, simply on the basis that it *is* a human being.

We obviously do the same with other species of animals. When I meet a cat or a gerbil, I treat it neither as an inanimate object like a sofa or a fridge or a rubber plant, nor as a person. (Not, of course, that I treat all inanimate objects the same, either. Sofas, fridges, and rubber plants are very different things, especially rubber plants.) You can play games with a gerbil. With a dog or a cat, you can even have a cuddle (in the reciprocal sense, I mean).<sup>xviii</sup> Neither is possible with a fridge. On the other hand, you would be wrong to expect a dog to play a good hand of contract bridge, or a cat to give you good investment advice, or a gerbil to read your mind, as a human being might. What you expect of an individual dog or individual cat will be different from what you expect of an individual gerbil, and both will be different from your expectations of inanimate objects. In each case, your expectations will be set by your background knowledge of the different kinds in question. Where dogs, gerbils, cats, and indeed all other kinds of living thing are concerned, this obviously means the different *species* in question. Why should it be otherwise with humans?

These remarks should enable us to see—*pace* the criterialists—how central it is, to our moral responses to the things and creatures around us, to assign them to species. It is species-assignments that enable us to answer what you might reasonably think is the most basic question of all about any thing that might be in front of us at any time: namely, what kind of thing it is. Until we know what species it is, we do not know how it may act, or how to treat it: what is good for it, how we might harm or hinder it, what counts as health for it, and so on. The nature of the creature determines what well-being is for it; and its nature is its species. To put it another way, the best way to answer “What is it?”, when this question is asked about any individual animal, is to give the name of its species. Or as we might ask moral individualists like McMahan: “If species-membership isn’t an intrinsic property, what *is*?” Singer, Rachels, and McMahan never tell us why *being human* cannot count as any creature’s “own particular characteristic”. It can’t be merely because *being human* is a group membership (as McMahan 2005 suggests). For there is nothing to stop a given property from being both a “group membership” and an “own particular characteristic”; that is how it is with *being a vertebrate*, for instance. To put it more formally, nothing stops a given property from being both a relational property, and also an intrinsic property.<sup>xix</sup> Indeed, if Kripke is right in his well-known thesis of the metaphysical necessity of origin, such properties must be very common.

This shows how humanism can be true without falling prey to a charge of arbitrary speciesism. To treat any species of creatures as persons is to treat them, not as successful passers of some test for the personal qualities, but on the assumption that they are individuals of a species the ideal for which is (roughly) the highest attainable development of those qualities. In this sense, references to species are an ineliminable part of our talk about persons.

In that sense, but not in some others. We are now in a position to respond to four familiar objections to “speciesism”. I consider these in section 4. The aim of my

response is not of course to defend speciesism, if that is defined as “*arbitrary* moral discrimination on the grounds of species”. But it is to defend humanism, the view that not all moral discrimination on the grounds of species *is* arbitrary.

#### 4. Four objections to speciesism

*a. All animals are equal.* Peter Singer (1993: 57) writes that “No matter what the nature of the being, the principle of equality requires that the suffering be counted equally with the like suffering of any other being”. Nothing I have said so far commits me to rejecting this application of the principle of equality. If I were to accept it, I would want to add, first, that plenty of other things besides suffering “count”, and secondly, that however alike human and animal *pain* (the physical sensation) may be, human and animal *suffering* (the state of mind) are mostly quite unlike. A cat can be tortured like Gloucester, but it cannot suffer like King Lear.<sup>xx</sup> And suffering matters more than pain.

But perhaps we should reject this application of equality anyway. Right now there is (quite likely) monstrous pain, and a fair amount of suffering too, on the plains of the Serengeti, where a wounded buffalo is being eaten alive by a pride of lions. Singer’s principle of equality requires us to say that our duty to prevent the buffalo’s pain and suffering in the teeth of the lions is directly proportional to the amount of pain and suffering involved. If any implication of any moral theory is absurd, this one is. There just isn’t any duty, not even a *prima facie* one, for us to act as Zoological Pain Police. (Don’t say: “We can’t intervene, because the lions might eat *us*”. Where is your moral courage? And don’t say: “But it causes less pain in nature overall if we don’t intervene in nature in this sort of way.” As many writers on the theological problem of evil are keen to remind us, most animal suffering is gratuitous. If lions can be caught and trained to jump through hoops, perhaps they can be caught and trained to kill their prey cleanly before eating it. So is anyone *trying* to train them? Of course they’re not.) The vast majority of the huge amount of animal pain and suffering that happens every day is simply none of our business. “It is not an accident or a limitation or a prejudice that we cannot care equally about all the suffering in the world: it is a condition of our existence and our sanity” (Williams 2006: 147).<sup>xxi</sup>

So it looks like Singer’s principle that all suffering counts for the same, no matter what the species involved, is just false. On the other hand, some weaker principle of this general sort seems clearly true: perhaps “Animal suffering is bad for the same sort of reasons as human suffering is bad”, or something like that. What also seems clearly true is that humans should do what they can not to generate animal suffering, and to prevent animals from being needlessly hurt or damaged *by humans* (and sometimes, though less often, by accident or by each other). This is a much weaker principle than Singer’s equality-of-suffering principle. But perhaps even this is strong enough to generate a good argument for vegetarianism. I suspect it all depends what counts as needless hurt or harm.

Hence “the limit of sentience” is not “the only defensible boundary of concern for the interests of others” (Singer 1993: 58). Certainly that limit marks one distinction between the ways we should take an interest in the welfare of other creatures. For example, hacking a frog to pieces is wrong in a way that hacking an

orange to pieces is not, partly because an orange certainly cannot feel pain, and a frog pretty definitely can. But only partly. Quite aside from the pain involved, destroying a living creature for fun is wrong in a way that destroying a piece of fruit for fun is not. (Hacking trees, or anaesthetised frogs, to pieces for fun is usually wrong too.<sup>xxii</sup>) Unless we think that nothing matters except pain and pleasure—and this is not, e.g., Singer’s position either (Singer 1993: xi)—there is no reason to think that the distinction between sentience and non-sentience is the only fundamental moral distinction we can make among beings; nor that it is the most important.

There are some interesting complications about the notion of sentience. If “sentience” means simply “capacity to feel (sensational) pain and pleasure”, then sentience does not fit the story that I have developed about how the personal properties come into being. The capacity to feel pain- and pleasure-sensations is certainly one of the capacities that we expect to find in any human person. But it is not, unlike rationality or intentionality, a capacity whose presence in any person is even partly constituted by others’ idealising assumption, expectation, or hope that it is there. Whether or not a being can feel pain does not depend on others’ adopting “participant attitudes” or taking the “intentional stance” towards him, in the way that (I have argued) his capacities for self-consciousness, rationality, and suffering do. I don’t think this shows that sentience is a counter-example to my thesis about how the person-constituting properties are developed. Rather, I think it shows that sentience is not one of the person-constituting properties. A creature which had the (other) person-constituting properties, but lacked the capacity for pain- and pleasure-sensations, would still undeniably be a person.<sup>xxiii</sup> A creature which had the capacity for pain and pleasure, but lacked the (other) person-constituting properties, surely would not be.

*b. Aliens, talking chimps, Locke’s parrot, and a possessed fridge.* The position that I am defending under the name of humanism says that, normally, to be a person it is *sufficient* to be a human being. Not *necessary*, because humanism does not imply that members of other species could not be drawn into the moral community of human persons. On the contrary, I can see at least three distinct ways in which this could happen. One is that we could discover a whole species—aliens, or angels, or what have you—which is like the human species in that its members are characteristically persons (in one sense or another). A second is that a species already known to us might, *en masse*, somehow change so as to become characteristically persons. Or, third, some particular individual creature—Locke’s parrot, say (Locke, *Essay* 2.27), or Washoe the chimp (Singer 1993: 111)<sup>xxiv</sup>, or some group of creatures, like Peter Carruthers’ Chimp Class of ’92<sup>xxv</sup>—might start to count as persons, perhaps e.g. by displaying the personal properties, even though this behaviour is otherwise completely untypical of the species in question. (Come to that, I suppose a possessed fridge might, in principle, start displaying the personal properties.)

In these special cases, we perhaps *will* have to appeal to something like the criterial view of personhood as our decision procedure for the vexed cases. Apparently we will have to ask whether this alien species, or this new class of highly-educated chimps, or this individual chimp, parrot, or spooky fridge, displays something like the criterialist’s personal properties, *before* we can decide whether to count these beings as persons.

Doesn't this show that my arguments against criterialism were mistaken? Not at all. It does show that something like criterialism may perhaps be our decision procedure if and when we encounter a being *that we do not already know how to classify as a person or not a person by way of its species*. (Perhaps; actually I doubt even this much is true, but at any rate it is the most that is true.) However, it would be utterly misleading to generalise from thought-experiments about these special and rare cases—almost all of which, to date, are imaginary—to alleged conclusions about the normal cases. (It would be as misleading to do that as it would be to generalise the justificatory/ rational structure of *every* belief we have about the external world from the justificatory/ rational structure of the external-world beliefs that we might form when we are thinking about the classic, and nearly always imaginary, sceptical predicaments.) Unfortunately, just this generalisation is a criterialist stock-in-trade. The most famous example of it is perhaps Mary Anne Warren's derivation of the personal properties (her list was quoted in section 1) from an imagined meeting with aliens. But even if Warren is right that this is the decision procedure that we would have to adopt with *aliens* to settle whether they were persons in some extraordinary imaginary scenario, nothing at all follows about our beliefs (either at the decision-procedure or at the criterion-of-rightness level) about whether any *humans* are persons in a multitude of ordinary real scenarios. For we have met humans before.

But anyway, I suspect Warren's description is much too black-and-white to get anywhere near characterising the decision procedure that would really be involved in such a meeting. (This is why I say, above, only that we would probably have to use *something like* the criterialist approach in alien encounters and similar.) For one thing, there is not the slightest reason to think that, in such a meeting, *yes* and *no* would be the only possible answers available to the question "Are these persons?". Even *Star Trek* yields a rather subtler answer—"It's life, Jim, but not as we know it"; the world, after all, is a complicated place, and it constantly throws us surprises. For another, our responses in such situations would be—or should I say *will* be?—based on all sorts of factors and reactions, most of them probably defying explicit articulation; on judgements about "forms of life", in fact. It would not and could not be based on some neatly finite check-list of diagnostic properties like Warren's. Looking beyond *Star Trek*, science fiction is full of wonderfully complex and subtle explorations of such encounters. The contrast with the mechanical simplicities of a criterialist list such as Warren suggests is instructive—and not *entirely* to be dismissed as unfair on the grounds that Warren is, after all, not writing science fiction.

*c. The two-species scenario.* A third interesting objection to speciesism is offered by Peter Carruthers:

Suppose it had been discovered that human beings in fact consist of two distinct species, otherwise hardly distinguishable from one another, the members of which cannot inter-breed. In these circumstances it would plainly be objectionable for the members of the majority species to attempt to withhold moral rights from the members of the minority, on the mere ground of difference of species. This, too, would be obvious speciesism. (Carruthers 1992 Ch 3)

Carruthers, I think, is quite right to claim that it would alter nothing morally speaking if we discovered that "the human species" was in fact two different

biological species—perhaps *homo neanderthalis* and *homo sapiens*—living together in a single moral community. (Modulo the point about not interbreeding, some anthropologists think that this has actually happened: Trinkaus and Shipman 1993.) Pace Carruthers, his argument does not show that there is no species-classification that bestows participation in the moral community on its members, and so that species-classifications are morally irrelevant. What his argument shows is that there may be *more than one* species-classification (or other kind-classification) that bestows participation in the primary moral constituency on its members. That is something that we should grant anyway given the possibility of other species whose members are characteristically persons; and it is not inconsistent with the claim that species-classifications are morally significant.<sup>xxvi</sup>

*d. Species and Darwinism.* A fourth objection to my argument is the claim that species don't exist: that as a matter of strict biology, there is no clear differentiation between humankind and "other species", only a differentiation between different parts of the same evolutionary family tree. Species therefore are populations, not Aristotelian (or Kripkean) natural kinds (Sober 1994). True, but irrelevant to my argument, which can work equally well with the population conception of species, or indeed with any conception at all that will allow "humanity" to count as the name of a kind of *some* kind. At least for creatures like us, whose temporal experience runs only over decades, hardly ever over centuries even, let alone the millions of years that speciation takes, it should not be a controversial thought that a population of creatures sharing a common genetic heritage, physiology, and ethology can sensibly be treated as a unitary grouping, as in fact common sense treats it. And that thought is all I am committed to meaning by "species".

So it looks like humanism can resist these four objections. And it looks like humanism can accept, indeed welcome, the thought that humans are not the only persons, the only members of the primary moral constituency—or at least, they might easily not be. What about the question on the other side: the question whether all human beings are persons? I turn to this question in the next and final section.

## **5. Are all human beings persons?**

Humanism as I have developed it here straightforwardly implies that all human beings are persons. Our treatment of any human being should be conditioned by the background of expectations, hopes, and aspirations that spell out what we know, from experience, humans in general can be. *Eudaimonia* in its broad outlines is the same for all human beings; and requires, as a general rule, that we must give all human beings the space to achieve *eudaimonia*—whether or not they predictably *will* achieve *eudaimonia*. To deny this space to any individual human being is to exclude that individual from the moral community of persons. And that is a serious injustice.<sup>xxvii</sup>

The very young are not excluded from the moral community of persons, just because they have not yet achieved many key forms of human *eudaimonia*; the mentally handicapped are not excluded from the moral community of persons, just because they cannot ever achieve many key forms of human *eudaimonia*. As I pointed out in section 1, *all* persons are incomplete and impure in their agency. No one ever



does everything she could with her own nature (partly, of course, because there are just too many things she could do). In one way or another, *all* of us are less than fully free to pursue happiness; everyone is wounded or damaged or limited or compulsive, or just plain ill, in some respect. That does not make it wrong to treat “normal” humans, whatever their age or IQ, with an interpretive charity, and with a kind of pity, that sees beyond their limitations, and leaves the door open, in each individual’s case, to what *any human* might become—even if we are well aware that *this person here* has no chance of becoming that.

[T]he kind of care that was needed to make us what we have in fact become, independent practical reasoners, had to be, if it was to be effective, unconditional care for the human being as such, whatever the outcome. And this is the kind of care that we in turn now owe or will owe. Of the brain-damaged, of those almost incapable of movement, of the autistic, of all such we have to say: This could have been us. (MacIntyre 1999: 100)

There is the possibility of deep moral concern for retarded people, in which they are seen as having, however incomprehensible we may find it, a human fate, as much as anyone else’s.... Someone may be very touched by the response of a severely retarded person to music; and there may be in that being touched an imaginative sense of shared humanity. (Diamond 1991: 55)

The terminally ill are not an exception to this rule either, just because they cannot any longer achieve many key forms of human *eudaimonia*. Perhaps it is true of them that they no longer have—to reuse Marquis’ phrase—“a future like ours”, and so will not achieve very much, or any, more *eudaimonia* before they die. Given that their futures have closed in on them in this radical way, you might call it “futile” to go on treating them with the full range of interpersonal hopes and expectations that you extend to any human, or to go on being as scrupulous about not killing them as you would be about anyone else. There again, you might also call it “respectful”; you might call it an expression of *pity*.

To pity someone in the sense I have in mind is to think, as MacIntyre puts it, “this could have been us”: this person too is part of human life and shares in human ideals, even if the idea of realising them in almost any way is quite hopeless for him/her. Such pity is a key part of our charity towards other people; of what is involved in recognising them as *people*, as fellow human beings.

(It is interestingly not something that follows in any very clear way from an Aristotelian virtue ethics where the focus is, eudaimonistically, on *maximal* forms of flourishing—on how much you as a being can develop, not on how many relatively undeveloped beings you can extend charity towards. Indeed it is a notorious fact about Aristotle that there is something rather ruthlessly bright-eye-and-gleaming-coat<sup>xxviii</sup> about his conception of *eudaimonia*, and that—not unconnectedly—he seems to have had little room in his ethics for the notion of pity.<sup>xxix</sup> Pity is, however, part of the sorts of virtue ethics that we may find in a Christian writer like Simone Weil. And that is the sort that I would defend, though I cannot defend it here.)

There are, as many people have emphasised, better and worse ways to die. And one of the worse ways to die is to have recognition of your humanity withdrawn

from you before it happens. Part of understanding why that is such a bad thing to have happen to you must, I think, depend on taking a different view from the criterialist's about the locus of the value of persons. For the criterialist, apparently, the value of persons consists in their instantiating the person-constituting properties. Aside from such instantiations, persons, at the most basic level, *have* no value (compare the familiar view in bioethics that being alive in itself is neither good nor bad); indeed there is a sense in which, aside from such instantiations, there *are* no persons. What such a view of the value of persons obscures from view is the notion of loving an individual in itself; it seems to be only the *properties* of an individual that a consistent criterialist can love—and then, only when they are duly manifested. But the idea that the object of love is always and necessarily some property-instantiation, common though it is, is a dogma. If property-instantiations can be objects of love (we may fairly ask), why not things in other metaphysical categories: individual human beings, for example?<sup>xxx</sup>

Alongside this dogma, perhaps there is another, related but slightly different, dogma at work in securing the foundations of criterialism. This is the dogma that whenever we answer the question “Why is X valuable?” by citing some valuable property V that X displays, what we must be really saying is always that it is (this instance of) V that is valuable, not X. But this too is a dogma, and unmasking it may be another way of setting ourselves free from the lures of criterialism. To cite V in explanation of the claim that X is valuable can, so to speak, be understood the other way around: you can take it that citing V is displaying a *symptom*, a *reminder*, of the value of X. When George Orwell, in a famous passage from his essay “A Hanging”, describes the growing fingernails and the refusal to walk through a puddle of the man who is about to be hanged, I take it that this latter thing is what he is doing.<sup>xxxi</sup> He is not saying that having growing fingernails is a value-giving property. Rather he is saying that *being human* is a value-giving property, the importance of which can be brought out for us by all sorts of exercises in the “assembling of reminders”—including such reminders as the growth of fingernails.

“But all the same, isn't there a profound difference between what we hope for *in general* for other people, and what we hope in the particular case of someone who is severely mentally handicapped?” There certainly is. There is always a distinction between the aspirations for other persons that are grounded in human *eudaimonia* in general, and the aspirations for specific other persons that are grounded in our particular knowledge of them as individuals. This distinction is just as applicable to the profoundly mentally handicapped as it is to anyone else. And it makes for the same commonalities, and the same differences, as with anyone else. If your friend is tone-deaf, there is no need to pretend that he will ever be Mozart; but that does not make it all right to take home all his CDs. If your friend loses both his legs in a skiing accident, you needn't spend the rest of his life pretending to him that he can still ski; but you miss something vital, too, if you bounce into his hospital ward straight after his double amputation with the words “So, you won't be wanting those ski-boots any more”. Many of the details of what counts as just and loving treatment of other persons are, certainly, set by the particular trajectories of their particular lives. But the most basic and elementary requirements of love and justice, is set by reference to the aspirations that arise from human *eudaimonia* in general. And this applies just as much to the mentally disabled as to anyone else.

In Nick Hornby's very funny novel about suicidal depression, *A Long Way Down*, Maureen, the mother of a severely mentally disabled child, says this:

This all began years ago, when I decided to decorate his bedroom. He was eight, and he still slept in a nursery—clowns on the curtains, bunny rabbits on the frieze round the wall, all the things I'd chosen when I was waiting for him and I didn't know what he was. And it was all peeling away... and I hadn't done anything about it because it made me think too much about all the things that weren't happening to him, all the ways he wasn't growing up. What was I going to replace the bunny rabbits with? He was eight, so perhaps trains and rocket ships and footballers were the right sort of thing for him—but of course, he didn't know what any of those things were... But there again, he didn't know what the rabbits were either, or the clowns. So what was I supposed to do? Everything was pretending, wasn't it? The only thing I could do that wasn't make-believe was paint the walls white, get a plain pair of curtains. That would be a way of [saying] that I knew he was a vegetable, a cabbage, and I wasn't trying to hide it. But then, where does it stop? Does that mean you can never buy him a T-shirt with a word on it, or a picture, because he'll never read, and he can't make sense of pictures? And who knows whether he even gets anything out of colours, or patterns? And it goes without saying that talking to him is ridiculous, and smiling at him, and kissing him on the head. Everything I do is pretending, so why not pretend properly?

In the end, I went for trains on the curtains, and your man from *Star Wars* on the lampshade. And soon after that, I started buying comics every now and then, just to see what a lad of his age might be reading and thinking about. And we started watching the Saturday morning television together, so I learned a little bit about pop singers he might like... thinking about these things helped me to see Matty, in a strange sort of way... I made up a son. (Nick Hornby, *A Long Way Down*, pp.118-9)

Of course there is—as Maureen herself recognises—something delusional, and something pathetic, in her treatment of Matty. For sure, her attitude to him is strained and exaggerated. But what we should not lose sight of is that it is a strained and exaggerated version of something morally indispensable: the kind of interpretive charity, the humanistic understanding and pity, that, I have argued here, we all need from everybody else all the time. In the words of an older writer: “Charity bears all things, believes all things, hopes all things, endures all things.”<sup>xxxiii</sup>

## Bibliography

- Bauby 1998: Jean-Dominique Bauby, *The Diving Bell and the Butterfly*. London: Fourth Estate, 1998.
- Carruthers 1992: Peter Carruthers, *The Animals Issue: Moral Theory in Practice*. Cambridge: Cambridge University Press.
- Chappell 1998: Timothy Chappell, *Understanding Human Goods*. Edinburgh: Edinburgh University Press.
- Chappell 2003: Timothy Chappell, "Absolutes and particulars", 95-117 in *Modern moral philosophy*, ed. A. O'Hear. Cambridge: Cambridge University Press.
- Davidson 1980: Donald Davidson, *Inquiries into truth and interpretation*. Oxford: Clarendon.
- Darwall 2006: Stephen Darwall, *The Second-Person Standpoint*. Cambridge, MA: Harvard University Press.
- Diamond 1991: Cora Diamond, „The importance of being human“, in David Cockburn, ed., *Human Beings*. Cambridge: Cambridge University Press.
- Engelhardt 1989: Tristram Engelhardt, "Ethical Issues in Aiding the Death of Young Children", in *Euthanasia: the moral issues*, edd. Robert M Baird and Stuart E Rosenbaum. Buffalo, N.Y. : Prometheus Books.
- Gaita 1991: Raimond Gaita, *Good and Evil: an absolute conception*. London: Routledge.
- Gaita 1998: Raimond Gaita, *A Common Humanity*. London: Routledge.
- Grau 2010: Christopher Grau, "Love and history", *Southern Journal of Philosophy* **48** (2010), 246-271.
- Grice 1957: H.P.Grice, "Meaning", *The Philosophical Review*, **66**: 377-388.
- Harris 1985: John Harris, *The Value of Life*. London: Routledge.
- Alison Hills, "Utilitarianism, contractualism and demandingness", *Philosophical Quarterly* **60** (April 2010), 225-242.
- Nick Hornby, *A Long Way Down*. London: Penguin, 2006.
- David Hume: *A Treatise of Human Nature* [1739].
- Kahane 2010: Guy Kahane, "Feeling pain for the very first time: the normative knowledge argument", *Philosophy and Phenomenological Research* **80.1** (2010), 20-49.
- Long and Sedley 1987: A.A. Long and D. Sedley, *The Hellenistic Philosophers*, Cambridge University Press, Cambridge, Vol.I, 67A
- Marquis 1997: Don Marquis, "An argument that abortion is wrong", in Hugh LaFollette, *Ethics in Practice*, First Edition (Oxford: Blackwell, 1997).
- MacIntyre 1999: Alasdair MacIntyre, *Dependent Rational Animals: why human beings need the virtues*. Chicago: Open Court, 1999.
- McMahan 2005: Jeff McMahan, "Our fellow creatures", *Journal of Ethics* **9** (2005), 353-380.
- Olson 2007: Eric Olson, Stanford Encyclopedia of Philosophy entry on "personal identity", <http://plato.stanford.edu/entries/identity-personal/>
- Parfit 1984: Derek Parfit, *Reasons and Persons*. Oxford: Oxford University Press.
- Patterson and Gordon 1993: "The Case for the Personhood of Gorillas", in Paola Cavalieri and Peter Singer (eds.), *The Great Ape Project* (New York: St. Martin's 1993), 58-77.
- Rachels 1990: James Rachels, *Created From Animals: The Moral Implications of Darwinism*. Oxford: Oxford University Press.
- Schapiro 1999: Tamar Schapiro, "What is a child?", *Ethics* **109.4**: 715-738.

- Singer 1983: Peter Singer, *The Expanding Circle*. Oxford: Oxford University Press.
- Singer 1991: Peter Singer, *A Companion to Ethics*. Oxford: Blackwell.
- Singer 1993: Peter Singer, *Practical Ethics*. Cambridge: Cambridge University Press.
- Singer 2002: Peter Singer, *Unsanctifying Human Life: essays on ethics*, ed. Helga Kuhse. Oxford: Blackwell.
- Sober 1994: Elliott Sober, *From a Biological Point of View*. Cambridge: Cambridge University Press.
- Strawson 1962: Peter Strawson, 'Freedom and Resentment', in his *Freedom and Resentment and other Essays*, London: Methuen, 1974.
- Teichman 1996: Jenny Teichman, *Social Ethics*. Oxford: Blackwell's.
- Tooley 1972: Michael Tooley, "Abortion and Infanticide", in Singer, ed., *Applied Ethics* (Oxford: OUP, 1987).
- Vlastos 1981: Gregory Vlastos, "The Individual as Object of Love in Plato", in his *Platonic Studies*, Princeton, NJ: Princeton University Press, 3–42, 2nd edn.
- Warren 1997: Mary Anne Warren, "On the Moral and Legal Status of Abortion", in H.Lafollette, ed., *Ethics in Practice* (Oxford: Blackwell, 1997).
- Williams 1985: Bernard Williams, *Ethics and the Limits of Philosophy*. London: Penguin.
- Williams 2006: Bernard Williams, *Philosophy as a humanistic discipline*. Cambridge: Cambridge University Press.
- Williams 2006a: Bernard Williams, "Hylomorphism", 218-230 in his *The Sense of The Past*. Cambridge: CUP.
- Ludwig Wittgenstein, *Philosophical Investigations*. Oxford: Blackwell, 1951.

## Notes

---

<sup>i</sup> Anscombe mistranslates *zur Seele* as “to a soul”; I have corrected this to “to *the* soul”.

<sup>ii</sup> Perhaps we could do ethics entirely or almost entirely without using the word or concept “person”: so numerous philosophers have at times suggested, including Jenny Teichman (1996), Bernard Williams (2006a), Christopher Coepe (2006), and myself (I have at least had this thought; I can’t now remember if it has ever seen the light of published day; something close to that thought is certainly at work in Chappell 1998, Chapter 4). However, I don’t pursue that idea here.

<sup>iii</sup> Criterialism, as I call it, is pretty much the same view as what McMahan 2005, following Rachels 1990, calls “moral individualism”. This is the view (Rachels 1990: 173) that “how an individual may be treated is determined, not by considering his group memberships, but by considering his own particular characteristics”.

<sup>iv</sup> Or something very close to criterialism. I need not commit myself yet on whether these authors are defending necessary-and-sufficient-conditions versions of a criterial view.

<sup>v</sup> Something like the criterial view pops up in many other places. For instance, it seems to be behind Olson 2007’s suggestion that “person” is a phase sortal: “Being a person may be only a temporary property of you, like being a philosopher”.

<sup>vi</sup> It might be suggested that (some instance of) Criterialism is a *bespoke* view of personhood, aimed only at defining what persons are for some specific purpose. Such a position is perfectly possible, of course, but I think it is very clearly *not* what is at issue in discussions like the four I cite here, all of which are aimed at completely general conclusions about “what persons are”: what they are in their essence, and for *any* possible purpose. (Thanks to an anonymous referee for discussion of this point.)

<sup>vii</sup> For more on personhood and language, see Patterson and Gordon 1993.

<sup>viii</sup> David Hume: “For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch *myself* at any time without a perception, and never can observe any thing but the perception. When my perceptions are removed for any time, as by sound sleep, so long am I insensible of *myself*, and may truly be said not to exist” (*Treatise* 1.6.3). Derek Parfit: “We could give a complete description of reality without claiming that persons exist” (Parfit 1984: 212). The Buddha: “There exists no individual, it is only a conventional name given to a set of elements” (quoted in Parfit 1984: 502).

<sup>ix</sup> Autism involves very various sorts of emotional dysfunction. It can involve *over*-emotionality and “inappropriate” emotional responses as well as a lack of emotional response. But that just underlines the absurdity of using “emotionality” as a necessary condition of personhood. For then autistics who are over-emotional will count as persons, autistics who are “cold” and “withdrawn” won’t.

<sup>x</sup> There is a surprising contrast between some recent philosophers’ willingness to grant moral standing to “potential persons” in the sense of metaphysically possible future human beings who admittedly do not exist right now, and their reluctance to do so to “potential persons” in the sense of young human beings who admittedly do exist right now.

<sup>xi</sup> “The wrongness of killing us is understood in terms of what killing does to us. Killing us imposes on us the misfortune of premature death. That misfortune underlies the wrongness... The misfortune of premature death consists in the loss to us of the future goods of consciousness... What makes my future valuable to me are those aspects of my future that I will (or would) value when I will (or would) experience them... What makes killing us wrong, in general, is that it deprives us of a future of value. Thus, killing someone is wrong, in general, when it deprives her of a future like ours.” (Marquis 1997: 95-96)

<sup>xii</sup> And it is nearly always *the* criterion, i.e. some *one* criterion, that we are talking about. (Warren’s list of six criteria is the only exception to this monism that I can think of.) The suspicion that the value of persons might be multiform not monistic, that there might be not just one simple and readily-articulable reason for valuing persons but *all sorts* of reasons, some of them far from simple or easy to put into words, is not at all a characteristic suspicion for criterialists.

<sup>xiii</sup> About the argument that I develop from this point on, about “how we treat” people, it is natural to ask: are these descriptive or normative claims? The answer to that, as quite often with Wittgensteinians like (I suppose) me, is: a bit of both. What is on offer is a phenomenology of (part of) the moral life. My claim is that the most basic and fundamental truths about this phenomenology makes certain options in thinking about ethics that might seem attractive and viable, such as criterialism, actually incoherent, because of its mismatch with what is basic and fundamental. Because of how things *are*, there are some things we *shouldn’t* think or do: both a normative and a descriptive thesis.

---

Another worry that might occur to some here: am I attacking criterialism because it is a *theory* of personhood? If so, isn't it hypocritical of me to go on to develop what is really just another theory of personhood? The answers are No and No. I'm not attacking criterialism because it's a *theory*, but because it's a *rotten* theory. A good theory (or if you dislike that word, a good philosophical account) of personhood will make good sense of our intuitions; this criterialism signally fails to do.

(My thanks to anonymous referees for pressing these questions.)

<sup>xiv</sup> The charming comparison is Singer's: Singer 1993: 89-90.

<sup>xv</sup> But when exactly? As part of the child's fourth birthday celebrations? On going to primary school? Would a Singerian society have a ritual for an induction into personhood?

<sup>xvi</sup> At least, the term "objective" is unfortunate if it implies, as I suspect it does in passages like the one from Strawson that I cite in the next Note, that there is something *non-objective* about participant reactive attitudes.

<sup>xvii</sup> "We look with an objective eye on the compulsive behaviour of the neurotic or the tiresome behaviour of a very young child, thinking in terms of treatment or training. But we sometimes look with something like the same eye on the behaviour of the normal and the mature. We have this resource [the objective attitude] and can sometimes use it. Being human we cannot, in the normal case, do this for very long, or altogether. But what is above all interesting is the tension there is, in us, between the participant attitude and the objective attitude. One is tempted to say: between our humanity and our intelligence. But to say this would be to distort both notions." (Strawson 1962: 66-67, with omissions)

<sup>xviii</sup> "One speaks and expresses emotions to and not simply at one's cat. Nonetheless, although we of course discipline them, we don't press claims against or hold our pets accountable in the same way we do with one another... Granted, when I look into my cat's eyes, I can't get over the feeling that he is looking also into mine in some personal way or shake the hope that he is seeing me in some way other than 'the guy who feeds me'. But I find it utterly impossible to sustain the thought that he can imaginatively enter into my point of view or acknowledge me as a being with an independent perspective." (Darwall 2006: 43)

Pets are an interesting case of something like a "pretend" interpersonal relationship, if "pretend" is the right thing to call it. We know that, unlike infants, our cats will *never* become responsive in the way that persons do. That doesn't stop us (or at least the dedicated pet-owners among us) from extending something interestingly like the constitutive prolepsis even to cats, hamsters, or stick-insects. Children, of course, do something similar with their dolls or other toys—and here "pretend" clearly is the right word. For a recent fashion to which similar comments apply see <http://www.nytimes.com/2009/07/26/magazine/26FOB-2DLove-t.html> . (Thanks to an anonymous referee for drawing this to my attention.)

<sup>xix</sup> "Intrinsic" in the sense of "essential", I mean; not in the sense of "non-relational", obviously.

<sup>xx</sup> Gaita 1991: 117: "An animal can suffer, but it cannot curse the day it was born. An animal can be afraid, but it cannot be ashamed of its fear and despite itself. An animal can be happy, but it cannot be joyous. An animal cannot give of its substance to certain pursuits and be admonished for doing so. One could go on almost indefinitely. The problems of life's meaning cannot arise for an animal. Only a being for whom life can be problematic can have a spiritual life, and therefore have a soul."

<sup>xxi</sup> For more examples applying this argument see Hills 2010: 235-7.

<sup>xxii</sup> So I say. In the case of trees, of course, there was at least one stern Victorian moralist who thought otherwise, namely W.E.Gladstone, whose hobby was chopping down trees.

<sup>xxiii</sup> For this speculation deployed for a different purpose, cp. Kahane 2010.

<sup>xxiv</sup> Cp. Gerald the gorilla in the well-known *Not the Nine O'Clock News* sketch:

<http://www.youtube.com/watch?v=MpbMm0433I>

<sup>xxv</sup> "Suppose that the experiments attempting to teach language to chimpanzees had been successful beyond their originators' wildest dreams. The apes in question gained a complete mastery of English within a few years, were able to attend school and later university, and made many close friendships with human beings... in these imaginary circumstances it would plainly be absurd to claim that the apes lacked moral standing, or had a moral importance that was lower than our own." (Carruthers 1992, Ch.3)

<sup>xxvi</sup> Carruthers' own view about animal rights is much closer to speciesism than to animal liberationism. He holds that there is an indirect argument, arising from contractualism, for granting all humans and no animals moral standing (for a brief exposition see his online paper "The animals issue" at <http://www.philosophy.umd.edu/Faculty/pcarruthers/The%20Animals%20Issue.pdf>). Since I think it is obviously true that (non-human) animals do have moral standing in their own right, based roughly on the wrongness of causing unnecessary suffering, damage, and destruction, it seems I am closer to animal liberationism than Carruthers is.

---

<sup>xxvii</sup> Jeff McMahan (2005: 366) writes that his “moral individualism” gives us reason to deny that it is a grave misfortune for a human individual who is disabled to have limited cognitive capacities. After all, it is not a misfortune for an individual animal to have capacities of roughly the same level; and according to moral individualism, it is only the individual creatures that we should be considering. So either both the human and the animal are unfortunate, or neither is.

This seems a very counter-intuitive consequence of “moral individualism”. Surely it matters how the creatures got into their states: at the very least, there must be a difference in fortune between a human individual who was born at a low cognitive level, and one who is at that level because of some horrendous accident? So far as I can see, McMahan must deny this. For presumably he thinks that my past stands to me as a relational characteristic, not an intrinsic one—and the whole point of “moral individualism” is that only “intrinsic characteristics” count morally.

<sup>xxviii</sup> Bernard Williams’ delightfully mordant phrase: Williams 1985: 46.

<sup>xxix</sup> The best presentation I know of the evidence for Aristotle’s blindness on the subject of pity is in Gaita 1991.

<sup>xxx</sup> For more on this point see Chappell 2003, Grau 2010, Vlastos 1981.

<sup>xxxi</sup> “It is curious, but till that moment I had never realized what it means to destroy a healthy, conscious man. When I saw the prisoner step aside to avoid the puddle, I saw the mystery, the unspeakable wrongness, of cutting a life short when it is in full tide. This man was not dying, he was alive just as we were alive. All the organs of his body were working—bowels digesting food, skin renewing itself, nails growing, tissues forming—all toiling away in solemn foolery. His nails would still be growing when he stood on the drop, when he was falling through the air with a tenth of a second to live. His eyes saw the yellow gravel and the grey walls, and his brain still remembered, foresaw, reasoned—reasoned even about puddles. He and we were a party of men walking together, seeing, hearing, feeling, understanding the same world; and in two minutes, with a sudden snap, one of us would be gone—one mind less, one world less.” (George Orwell, “A Hanging”, *Adelphi Magazine* August 1931)

<sup>xxxii</sup> My thanks for their comments and encouragement to a large number of people, including Julia Annas, Alex Barber, Chris Belshaw, David Benatar, Peter Cave, Stephen Clark, John Cottingham, Christopher Coope, Christopher Grau, John Haldane, Rosalind Hursthouse, Tim Mulgan, David Oderberg, Tim O’Hagan, Charles Taliaferro, Chris Tollefsen, and Garrath Williams; an audience in Leeds at the Society for Applied Philosophy in 2009; and three anonymous referees for this journal.