

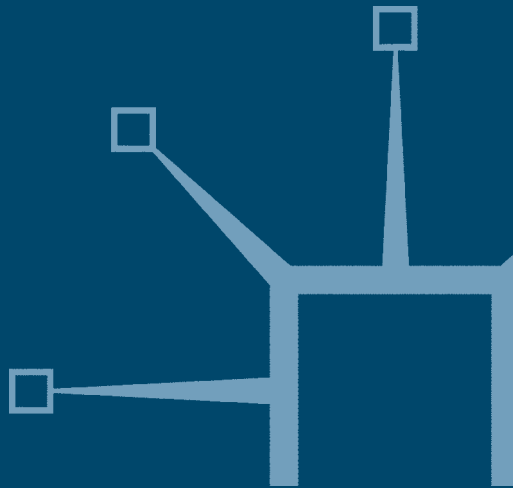
palgrave
macmillan

Animal Rights

Moral Theory and Practice

Second Edition

Mark Rowlands



Animal Rights

Also by Mark Rowlands

SUPERVENIENCE AND MATERIALISM (1995)

THE BODY IN MIND (1999)

THE NATURE OF CONSCIOUSNESS (2001)

EXTERNALISM (2003)

BODY LANGUAGE (2006)

ANIMAL RIGHTS (1998)

THE ENVIRONMENTAL CRISIS (2000)

ANIMALS LIKE US (2002)

THE PHILOSOPHER AT THE END OF THE UNIVERSE (2003)

EVERYTHING I KNOW I LEARNED FROM TV (2005)

FAME (2008)

THE PHILOSOPHER AND THE WOLF (2008)

Animal Rights

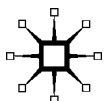
Moral Theory and Practice

Second Edition

Mark Rowlands

University of Miami, Florida, USA

palgrave
macmillan



© Mark Rowlands 1998, 2009

All rights reserved. No reproduction, copy or transmission of this publication may be made without written permission.

No portion of this publication may be reproduced, copied or transmitted save with written permission or in accordance with the provisions of the Copyright, Designs and Patents Act 1988, or under the terms of any licence permitting limited copying issued by the Copyright Licensing Agency, Saffron House, 6-10 Kirby Street, London EC1N 8TS.

Any person who does any unauthorized act in relation to this publication may be liable to criminal prosecution and civil claims for damages.

The author has asserted his right to be identified as the author of this work in accordance with the Copyright, Designs and Patents Act 1988.

First edition published 1998

Second Revised Edition published 2009 by

PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries.

ISBN-13: 978-0-230-21944-1 hardback

ISBN-13: 978-0-230-21945-8 paperback

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

A catalogue record for this book is available from the British Library.

A catalog record for this book is available from the Library of Congress.

10 9 8 7 6 5 4 3 2 1

18 17 16 15 14 13 12 11 10 09

Printed and bound in Great Britain by

CPI Antony Rowe, Chippenham and Eastbourne

Contents

1	Animal Rights and Moral Theories	1
2	Arguing for One's Species	8
1	The Independence Day scenario	8
2	The opening gambit: how to argue (morally) with aliens	11
3	The alien response	16
4	Thou art the man	26
5	Who speaks for wolf?	29
3	Utilitarianism and Animals: Peter Singer's Case for Animal Liberation	31
1	Utilitarianism I: definition of utility	32
2	Utilitarianism II: maximizing utility	36
3	Problems with utilitarianism	42
4	Utilitarianism and justice	48
5	Rule utilitarianism	51
6	Utilitarianism and animals	53
4	Tom Regan: Animal Rights as Natural Rights	58
1	Subjects-of-a-life	59
2	Inherent value	61
3	Inherent value as a theoretical postulate	63
4	The respect principle	66
5	The harm principle	67
6	Moral rights	68
7	The miniride principle	71

8	The worse-off principle	74
9	Regan on vegetarianism	77
10	Inherent value is mysterious	86
11	Inherent value is <i>ad hoc</i>	89
12	Inherent value is unnecessary	93
5	Virtue Ethics and Animals	98
1	The concept of virtue?	98
2	Virtue ethics and animals: Scruton versus Hursthouse	100
3	The virtue of mercy	113
4	Conclusion	117
6	Contractarianism and Animal Rights	118
1	Contractarianism and moral status	118
2	Two forms of contractarianism	123
3	Contractarianism and animal rights: the orthodox view	129
4	Rawls and contractarianism	131
5	The original position revisited	142
6	The original position and animal rights	147
7	Hobbesian remnants	152
8	Other objections	158
9	Contractarianism and vegetarianism	162
10	Conclusion	174
7	Animal Minds	176
1	Introduction: morality and mentality	176
2	HOT and the rejection of animal consciousness	181
3	The problem of regress	183
4	The holism of the mental	187
5	An unsuccessful refutation	193
6	Attributing content	196
7	Belief and truth	209

8 How animals can refer	211
9 Conclusion	217
<i>Notes</i>	219
<i>Bibliography</i>	227
<i>Index</i>	231

For Emma

1

Animal Rights and Moral Theories

The contemporary philosophical arm of the animal rights or liberation movement effectively began in 1975 with Peter Singer's book *Animal Liberation*.¹ In this work, and in subsequent development of its ideas,² Singer argues that the moral theory known as *utilitarianism* can be used to justify and defend the moral claims of non-human animals. According to utilitarianism, a morally good action is one which promotes or produces the greatest amount of pleasure, happiness, or satisfaction of desires, and Singer argues, quite forcibly, that such promotion requires abandoning such practices as animal husbandry, and experimentation upon animals for scientific or commercial purposes. Singer's case for animal liberation, then, is anchored in his adoption of a utilitarian moral theory.

In 1983, Tom Regan published his important work *The Case for Animal Rights*.³ Rejecting Singer's utilitarianism, Regan argued that many sorts of non-human animals possess moral rights because they possess what he referred to as *inherent value*. In virtue of this, Regan argued, we are morally obligated to treat them in ways that respect this value. And, for Regan as for Singer, this requires us to abandon such practices as animal husbandry, vivisection, and so on. Inherent value for Regan is an objective property, and whether or not an individual possesses it does not in any way depend on whether he, she, or it is valued by others. Whether or not a person possesses inherent value depends only on their *nature* as the type of thing they are. And this places Regan, at least in one

important respect, in the tradition constituted by the doctrine of *natural rights*. Or, at least, it makes him an important intellectual inheritor of this doctrine.

To claim that Singer's *Animal Liberation* and Regan's *The Case for Animal Rights* are the two seminal works of the contemporary philosophical literature on animals would not, I think, be inaccurate. This, of course, is not to deny that there have been other important contributions. Philosophical analysis of the moral issues raised by non-human animals is a burgeoning field, and some of the contributions to this field have been quite excellent.⁴ However, I think it is true to say that, in terms of the widespread circulation and recognition of their work, by both philosophers and non-philosophers, by both friends of animal liberation and its foes, the work of Singer and Regan has been the most influential. And this means that any attempt to adjudicate the moral claims of animals, or on the moral issues raised by animals, must, effectively, define itself in relation to the work of Singer and Regan. This book is no exception. I discuss and critique Singer's utilitarian defence of the claims of animals in Chapter 3, and Regan's rights-based defence of animals in Chapter 4.

A case for the moral claims of animals *solely* defined by its relation to the work of Singer and Regan, however, would be incomplete in at least two ways. First of all, since the first edition of this book came out in 1998, there has been a welcome re-emergence of the tradition of *virtue ethics*.⁵ This second edition, therefore, incorporates a new chapter – Chapter 5 – which develops a virtue ethical underpinning for understanding our obligations to animals.

Secondly, one of the primary concerns of this book – both the first edition and this new edition – is to provide a contractarian or contractualist case for the moral claims of animals. For the philosophical defender of animals, this is an important task. *Contractarianism* or *contractualism* is, historically speaking, at least as important as the utilitarianism employed by Singer or the natural rights doctrine wielded to great effect by Regan. Therefore, if a philosophical defence of the moral claims of animals is to be

secure, it must be shown that such claims are derivable not only from a utilitarian approach, and not only from a natural rights approach, and not only from a virtue ethical standpoint, but also from a contractarian approach. This, in many respects, is the most difficult task: contractarian approaches are widely thought to be inimical to the moral claims of animals. Such approaches view one's moral rights and duties as deriving from the terms of an agreement reached by contractors in a hypothetical bargaining situation. Non-human animals, being non-rational, cannot plausibly be regarded as contractors in such a situation. And, therefore, it is widely assumed, non-human animals cannot be the bearers of moral rights or entitlements, and, conversely, we have no duties towards them.

This view of contractarianism provides common ground for both opponents and proponents of animal rights. Peter Carruthers, an opponent, has developed this argument quite forcefully in his book, *The Animals Issue*.⁶ Carruthers accepts that contractarianism provides the most adequate basis for a moral theory, and this provides the framework for his case against non-humans. Because animals are not rational agents of the sort who can plausibly be regarded as framers of a contract, they lack moral status. On the other hand, Tom Regan, perhaps the staunchest supporter of the concept of animal rights, and certainly one of its most important intellectual progenitors, has attacked contractarian moral theories on precisely these grounds.⁷ Regan, too, believes that contractarianism is incompatible with the attribution, to non-human animals, of moral rights.

The view that contractarianism is incompatible with animal rights, then, is both widespread and tenacious. In this book, however, I shall argue that this view is simply false. Firstly, contractarian moral theories are certainly compatible with possession of moral rights by non-human animals, and by non-rational humans. Secondly, contractarianism, properly understood, provides (perhaps) the most satisfactory theoretical basis for the attribution of moral rights to non-human and non-rational individuals. Far from

being a thorn in the side of the friend of animal rights, then, contractarianism is, in fact, possibly her greatest ally.

I say that contractarian approaches are widely assumed to be inimical to the moral claims of animals. I wish I could say that since the first edition of this book, and also its more practically oriented sequel,⁸ they are not *as* widely assumed to be inimical to these claims as they once were. But I'm really not sure about that. I am sure, however, that my original contractarian defence of the moral claims of animals has engendered a number of misunderstandings and superficial rebuttals. These misunderstandings and responses all seem to turn on a failure to properly distinguish two quite different forms of contractarian theory.

On the one hand, there is the form of contractarianism which derives, in a fairly direct way, from Hobbes. This form emphasizes the benefits, in terms of protection of life, limb, and property, which a contract affords. We might refer to this as *Hobbesian contractarianism*, and interpreted in this way, the contract is an essentially prudential device, its purpose consisting in the security it provides.⁹ It should be clear that this view is going to have a very difficult time accounting for our moral commitments to certain sorts of human beings, let alone non-human animals. If the point of the contract derives from the protection it affords us, and if we only need protection from those individuals who are a threat, or possible threat, to us, then, all other things being equal, there is simply no point in contracting with those individuals who are sufficiently weaker than oneself that they pose no real threat. For the Hobbesian contractarian, morality reduces to rational self-interest. And rational self-interest will extend the scope of one's contractual commitments only as far as those individuals who in some way constitute a threat, including a threat by proxy, or to those individuals with whom contracting might yield some advantage. It is the Hobbesian interpretation of contractarianism that is the primary motivation for the supposition that contractarian moral theory is incompatible with the attribution of moral rights to non-humans. Animals, in general, pose very little threat to us.

And, more importantly, not being rational agents, they cannot coherently be regarded as contractors. Therefore, we cannot contract with them, and have nothing to gain from attempting to do so. Therefore, we should not contract with them. And, therefore, they lack moral status.

There is, however, another, very different, way of developing the idea of the contract, a way that has its roots in the work of Kant, and receives its most influential recent formulation in the work of John Rawls.¹⁰ We can, again following Kymlicka, refer to this interpretation of the contract idea as *Kantian contractarianism*. The central concept underpinning this interpretation is that of the moral equality of all individuals, and the resulting ideal of *impartiality* as constitutive of moral deliberation. In Rawls's work, impartiality in moral deliberation is safeguarded by an imaginative, and purely heuristic, device known as the *original position*. The contractors in the original position find themselves behind a *veil of ignorance*. That is, each contractor has no knowledge of his or her natural talents and characteristics – his or her intelligence, gender, physical appearance, athletic aptitudes, and so on. Nor do they know their position in society. In fact, each contractor does not even know his or her conception of the good, the things they value, the things they despise, and so on. From behind this veil of ignorance, then, each contractor is, in effect, forced to be impartial in their deliberations. One can be partial towards oneself only if one knows who, and what, one is. This book argues that *Kantian contractarianism* provides a theoretically viable framework for the attribution of moral rights to non-human animals. In fact, I shall argue that the framework it provides in this regard is demonstrably superior to those of its traditional utilitarian and rights-oriented competitors. Kantian contractarianism, then, provides the first plank in the central argument of this book.

Logically these theories are quite different, both in their form and their consequences. Historically, they have been run together. Properly separating these different versions of contractarian theory is essential to a convincing contractarian defence of animals. The result of this labour is a greatly expanded Chapter 6.

There is one other addition to the first edition: the final chapter in which I discuss animal minds – attributing mental states to animals. When I wrote the first edition, I assumed – I thought safely – that no sane person would deny that animals were what is known as *phenomenally* conscious. That is, they have sensations and experiences, and when they do, things feel or seem a certain way to them. Equivalently: I assumed that there is *something that it is like* to be a non-human animal, just as there is something that it is like to be a human animal. If there were any interesting philosophical disputes, I thought, they would pertain to more complex mental states – thoughts, beliefs, and the other of the so-called propositional attitudes. Accordingly, the 1998 version of the final chapter spent most of its time trying to disarm a well-known argument, associated with Donald Davidson (and also Stephen Stich), which suggests that attributing such mental states to non-human animals is problematic. However, incredibly I think, since 1998 some have questioned whether animals are even conscious. The motivation for this is the higher-order thought (or HOT) model of consciousness, according to which for a mental state to be conscious requires the subject of that state – the creature that has it – to have a thought to the effect that it is in this state. Animals can't do this, it is claimed; therefore they are not conscious. This new version of the final chapter, therefore, also contains an extended discussion of this view of consciousness.

One of the defining characteristics of growing older is, I think, a certain loss of confidence. That is as it should be. Certainty is the preserve, and perhaps the prerogative, of the young. And so, ten years on, I'm not as sure of things as I used to be. The standard arguments against utilitarianism, which I shall rehearse in Chapter 3, don't now seem to be as compelling as they used to be. Tom Regan's rights-based defence of animal rights doesn't, now, seem to me to be as metaphysically outrageous as it used to. I'm no longer convinced that, from the perspective of moral theory, contractarianism is the only game in town.¹¹ I'm naturally attracted to virtue ethics, but can't shake the suspicion that it doesn't really

bring anything new to the table. I have deep reservations about all of the moral theories I am going to discuss in this book; and even deeper reservations about the possibility of decisively adjudicating between them. But one thing of which I'm still pretty certain is this: collectively, these theories pretty much consume all the available options; they exhaust the possible theoretical moves in moral space. When you put them together, in other words, they *do* constitute the only game in town. So, if it can be shown that each of these theories can be used to underwrite the significant moral claims of animals, this is as decisive a demonstration as it is possible to get in moral theory that animals do, in fact, have significant moral status. The goal of this book is to show that this is indeed the case.

2

Arguing for One's Species

Who speaks for wolf?

Iroquois invitation

The rich man had exceeding many flocks and herds. But the poor man had nothing save one little ewe lamb, which he had bought and nourished up: and it grew up together with him, and with his children; it did eat of his own meat, and drink of his own cup, and lay in his bosom, and was unto him as a daughter. And there came a traveller unto the rich man, and he spared to take of his own flock and of his own herd to dress for the wayfaring man that was come unto him; but took the poor man's lamb, and dressed it for the man that was come unto him.' And David's anger was greatly kindled against the man; and he said to Nathan ... 'the man that hath done this thing shall surely die ... because he had no pity.' And Nathan said to David, 'Thou art the man.' (Samuel 12: 2-7)

1 The Independence Day scenario

Suppose the earth were to be invaded by a species of powerful aliens. These aliens, for reasons which will no doubt become almost immediately clear in the embarrassingly unsubtle and thinly veiled story to follow, we can call *namuhs*. The intentions of these creatures are in no way benevolent, at least not towards us. In fact, they make it very clear that their primary purpose in

invading this planet has to do with food. They plan to cultivate and farm the earth's planetary fauna, which has a galaxy-wide reputation as being of excellent nutritional value. In particular, at the top of their list of desirable food species is the human race, the meat from which is regarded as one of the tastiest in this part of the galaxy. We might call this the *Independence Day* scenario, after the rather successful film with a similar story line. Actually, in the film, it is not clear if the aliens intend to eat humans; simply killing us seems to cohere better with the overall story line. (Therefore, it might have been better to have called it the *lifeforce* scenario, after Tobe Hooper's significantly lower budget, less successful, but still cult 1983 movie, although even here the aliens didn't plan to eat us exactly, but only appropriate our lifeforce. In the end, the wider circulation of the former won out, and the *Independence Day* scenario it is.)¹ Let's suppose that in our *Independence Day* scenario, the plan of the aliens is to engage in what we might call *human husbandry*: they plan to raise, kill, and eat us. This sort of scenario is, of course, a common science-fiction theme. Let us suppose also that there are certain features which the aliens possess.

Firstly, their intelligence is vastly superior to ours. In fact, so much is this so that they regard us in much the same way as we regard other higher mammals. They think of us in much the same way as we think of dogs and cats, pigs and cows, sheep and poultry. And let us suppose that they are largely correct in this estimation of our relative intelligence. That is, the difference in intelligence between us and, say, dogs, is roughly the same as the difference in intelligence between the aliens and us. This difference in intelligence results in them having a vastly superior technology which allows them to subdue us quite easily.

Secondly, although they take great pleasure in eating meat, especially human meat, they do not require meat in order to survive. They can survive, indeed flourish, on a purely vegetable diet. Their roving the galaxy in search of fresh supplies of meat stems only from the fact that they enjoy eating meat much more than vegetables, and from the fact that meat eating is their traditional

diet, adopted by their fathers and their fathers before them, and so on.

Thirdly, with regard to interactions between themselves, the aliens adhere to a strict moral code. In fact, they have evolved a democratic culture and society in many ways similar to our own. The conceptual centrepiece of this democratic culture is a principle of equality: roughly the idea that each of the namuh is to be treated with equal consideration and respect. Of course, it is not clear, even to themselves, what treating namuhs with equal consideration and respect amounts to. And their philosophers spend a good deal of time arguing over this. Some, regarded by the aliens as being on the left wing of their political culture, emphasize equality of welfare: at the very minimum, the welfare needs of everyone are to be taken care of. Others, of a more right-wing alien political persuasion, put much more emphasis on equality of opportunity: everyone is to be given equal opportunity to make whatever they can of their lives. Nevertheless, in their culture, while there are different and competing interpretations of the idea of treating everyone with equal consideration and respect, the aliens all agree that, whatever the best interpretation of this idea turns out to be, everyone should be treated with equal consideration and respect. Furthermore, they recognize that this principle of equality is not a *description* of an actual equality that holds between them. They recognize that some of them are more intelligent than others, that some of them are physically more powerful than others, that some of them have skills and aptitudes that others do not possess, and so on. So, they recognize that if the principle were to be interpreted as a description of an actual equality that existed between each of them, then the principle would almost certainly be false. But this is not, in fact, the case. The principle is not a description of an actual equality, but a *prescription* for how each of them is to be treated. The principle claims that each alien is to be treated with equal consideration and respect, whatever their level of intelligence, whatever their physical strength, whatever their skills and aptitudes might be. This is, in many ways, the fundamental moral principle of their society; the

principle from which all others stem. And the aliens take it very seriously, or at least *profess to do so*.

Finally, their adherence to a moral code means that the aliens are not morally blind. They recognize moral considerations and moral arguments when they see them, and they can be swayed by these considerations and arguments.

Let us suppose, then, that the alien invasion of earth is proceeding apace, and more and more humans are finding themselves on what are essentially factory farms. Unfortunately, in our Independence Day scenario there is no heroic American president (played heroically by Bill Pullman) to save humanity. Nor is there a feisty US air-force pilot (played feistily by Will Smith) to force the aliens to think again. In fact, the fate of humanity lies in the hands of a few philosophers who have hitherto escaped capture. Curtains for humanity, one might think. However, the philosophers, heartened by the fact that the aliens recognize moral considerations and arguments, and can be swayed by these considerations and arguments, decide to try and convince the aliens that what they are doing is wrong. If you were one of these philosophers, how would you go about this? How would you attempt to convince the aliens that the practice of human husbandry was wrong?

2 The opening gambit: how to argue (morally) with aliens

This is how to argue with aliens. First of all, we examine in more detail their conception of morality in general, and the principle of equality in particular. The principle of equality, of course, is a moral or ethical principle; that is, it is a principle which states what sort of behaviour is required if the demands of morality are to be met. In examining their conception of morality, however, we find that this moral principle rests on a further meta-ethical principle. A meta-ethical principle, in this sense, is one which is not a moral principle as such, but which provides a justification for a particular

moral principle. So, when we examine the alien morality, we find, first, a moral principle:

Each namuh should be treated with equal consideration and respect (whatever that amounts to).

and, secondly, a meta-ethical principle:

No moral difference without a relevant natural difference.

A natural difference, here, is simply a non-moral difference. The second principle, then, claims that there can be no moral difference without a non-moral difference. The moral principle, the principle of equality, holds because, or in virtue, of the meta-ethical principle.

The meta-ethical principle provides, for the aliens, a constraint on the way they think about morality and the way they use moral language. And this principle applies to all things that can be the subject of moral evaluation: persons, actions, events, rules, institutions, and so on. Consider two aliens whom we can call 'Worf' and 'Schworf'. Suppose Worf and Schworf are very similar. In fact, they have pretty much the same qualities and features. Both are honest, courageous, and benevolent (at least by namuh lights), and both tend to be a little rash and belligerent. In short, with respect to any features that might conceivably go into making a moral evaluation of them, Worf and Schworf are identical. Then, by the meta-ethical principle that there is no moral difference without a relevant natural difference, Worf and Schworf must be given the same moral evaluation. That is, either both must be good or neither is. Given that there is no difference in their relevant natural properties, it would make no sense to say that Worf is good but Schworf is not, or that Schworf is good but Worf not. A difference in moral evaluation would be justifiable only if there is a relevant natural difference between the two, and, *ex hypothesi*, in our example, there is not. The same sorts of considerations apply to all other things

that can be the subject of moral evaluation. Thus, if both Worf and Schworf help distinct elderly female namuh across the road, and assuming there is no natural difference between their actions (e.g., both elderly female namuhs want to cross the road, etc.) then if Worf's action is good, Schworf's must be good also. Conversely, if Schworf's action is bad, Worf's must be also. A difference in the moral evaluation one makes of two actions can only be justified if there is a relevant natural difference between those actions and, *ex hypothesi*, in our case there is not.

Notice that there is nothing in the claim that there can be no moral difference without a relevant natural difference which requires that moral evaluations are logically entailed by natural properties. The claim, for example, is not that because Worf has certain natural properties, this entails that he is good. Rather the claim is that if Worf has certain natural properties and he is also good, then any other individual – Schworf, or whoever – who has precisely the same natural properties as Worf, must also be good. Unless there is a relevant natural difference between two individuals, both must be given the same moral evaluation: either both are good, or neither are.

In the interests of precision, and when the fate of the human race lies in your hands it might pay to be precise, we might formulate the meta-ethical principle in the following way:

- (S) For the set of all moral properties M, and the set of all natural properties N, necessarily, for any objects x and y, if x and y share all properties in N then x and y share all properties in M – that is, indiscernibility with respect to N entails indiscernibility with respect to M.

Once again, a natural property is to be understood simply as a non-moral property. (S) states that any two objects – where an object can be understood broadly as including persons, actions, events, institutions, and the like – that are identical with respect to the natural properties they possess – must also be identical with respect to the

moral properties they possess. This is often put by saying that moral properties *supervene* on natural ones. And namuh moral philosophy is based on the idea that moral properties are supervenient upon natural ones in this sense.

The fundamental ethical (as opposed to meta-ethical) principle held by the namuhs is that all namuhs – whatever their intellect, strength, skill, and assorted aptitudes – should be treated with equal consideration and respect. As we have seen, namuhs disagree about what exactly treating individuals with equal consideration and respect amounts to, but they still agree on the fundamental idea. The reason they believe this can now be explained by the meta-ethical principle that moral properties *supervene* on natural ones. If we are to justify treating one individual differently from how we treat another individual, we must be able to cite a relevant natural difference between those two individuals. Sometimes differential treatment of two individuals, or between groups of individuals, can be justified in this way. Suppose, for example, that a certain segment of the namuh population is, because of slight genetic differences from the remainder of the population, susceptible to a certain disease against which the remainder of the population is immune. Suppose, further, that this disease can be prevented or controlled through an early screening procedure. This difference between the two sections of the population is sufficient to justify a certain sort of differential treatment. In particular, the namuhs could justifiably deny the screening procedure to members of one section of the population while making it available to members of the other. In this case, there does seem to be a relevant natural difference between the members of each section; a natural difference which justifies, or could justify, this sort of differential treatment. When the namuhs deny certain members of their population access to the screening procedure, it is not true that they are thereby failing to treat them with equal respect. Since those members of the population are not susceptible to the disease, treating them with respect does not require making the screening procedure available to them.

On the other hand, according to the namuhs, the fundamental idea of treating individuals with consideration and respect is

of a different order. If we are to treat one individual with consideration and respect, then we can only, justifiably, not treat another individual in the same way if there is a relevant natural difference between the two. In other words, the moral entitlements of individuals – what an individual is morally entitled to – supervene on natural properties. Therefore, there can be no difference in the moral entitlements possessed by two individuals unless there is some relevant natural difference between them. And, according to the namuhs, with respect to the entitlement to be treated with consideration and respect, there is no relevant natural difference between individual namuhs. While individual namuhs may differ with respect to their intelligence, physical strength, skills, aptitudes, and so on, none of these, according to accepted namuh morality, constitute relevant natural differences that could justify treating one sort of namuh with consideration and respect while denying that treatment to another. Therefore, because there are no relevant natural differences between namuh, all must be treated with equal consideration and respect. Thus speaks the namuh moral law.

So far, we have identified the fundamental ethical principle of namuh morality, and the fundamental meta-ethical principle from which the former derives. We have seen also how the ethical principle derives from the meta-ethical principle. This gives us a bridgehead; a platform from which to engage the namuhs in moral dialogue.

One way of arguing in ethics goes like this. You identify what your opponent believes, call it X, and then try to show that if she believes X, then she is also logically committed to another claim Y. And Y is hopefully the claim that you yourself endorse. This, for me, is the best way to engage in ethical argument. The alternative is to try to derive your own view from *first principles*. First principles are principles which, so it is thought, nobody could reasonably contest. The basic problem with invoking principles that nobody could reasonably contest is that such principles are typically contested, and often quite reasonably. Now, when the future of humanity is at stake, it is probably best not to rely on first principles since it would be somewhat disappointing, to say the least, to find that one's alien

interrogators do not share your apparently incontestable principles. Far better, far safer, to rely on principles you know your interrogators share. Whether you agree with those principles or not is not really relevant. What is relevant is that the namuhs believe them. That is all we need.

In other words, the strategy we can now adopt is this: we have identified the fundamental ethical and meta-ethical principles on which namuh morality is based. Now what we need to show is that these principles commit the namuhs to the claim that the practice of human husbandry is morally wrong.

Given the adherence of the namuhs to the meta-ethical principle that there can be no moral difference – including a difference in one’s moral entitlements – without a relevant natural difference, the following strategy is clearly the one to adopt. We try to show that there is no relevant difference between us and the namuhs. While there are, of course, clear differences between us and them, none of these differences are morally relevant ones. Therefore, if namuhs are morally entitled to be treated with consideration and respect, so too are humans. If we can show this, then we have shown that the namuhs, by their own moral principles, are committed to treating humans with consideration and respect equal to that which they show to their own.

This task, however, can only be achieved by finally engaging the namuhs in the dispute. We can only show that there is no morally relevant difference between us and them if we consider, assess, and finally demolish whatever proposals they put forward for the relevant difference, or differences, between them and us.

3 The alien response

The namuhs are likely to have several proposals for what constitutes the morally relevant difference between them and us. The first of these is the most obvious.

Species membership

According to this suggestion, the relevant natural difference between namuhs and humans, the difference which accounts for the fact that the former are entitled to be treated with consideration and respect, while the latter are not, is simply that the former but not the latter, are members of the namuh species, or the namuh *race*, as they prefer it to be called. Namuhs are all entitled to be treated with (equal) consideration and respect simply because they are namuh. Humans do not deserve to be so treated because they are not namuh. Thus, species membership provides the criterion of what is known as *moral considerability*, where an individual is morally considerable if it is morally entitled to consideration and respect. Even with our vastly inferior human intellectual capacities, however, it is easy to see that there are formidable problems with this suggestion. We might put the matter to our namuh disputants in this way. Suppose it were to be discovered that a certain substantial proportion of the namuh population were not really namuhs at all. Whether or not one is a namuh is, after all, a question of one's genetic structure, since the category of a species is ultimately a genetic one. Let us suppose that these deviant inhabitants of the planet *htrae* (you've guessed it! and pronounced 'hut-ray-eh', by the way) are of extra-htraen origin, coming from another planet in the htraen solar system. However, by the sort of logically possible coincidence for which philosophers will be eternally grateful, these non-htraens exactly resemble the namuhs. There are no phenotypic differences between the two, and while the genetic differences are significant enough to constitute the non-htraens as a distinct species from the namuhs, the namuhs, for religious reasons, banned any form of genetic research early in their development, and so are unable to detect these differences. As a result, the non-htraens became useful and much loved members of namuh society. Neither the namuhs nor the non-htraens were aware of the fact that they belonged to different species. Intermarriage was common between the two groups,

though, of course, no one realized that this is what it was. And no one seemed to be perturbed by the fact that many of the marriages on htrae were childless, or resulted in sterile children. Such a situation was thought of as quite natural. Now, the question to put to the namuhs is quite obviously this. If such a state of affairs were in fact to be the case, what would you say about the moral status of the non-htraens? Do they in fact have no moral entitlements, simply because of these genetic differences? Would one be willing to deny one's husband, or wife or (albeit sterile) children any moral entitlements, simply because of this genetic difference? Indeed, the point can be pushed further. How do you know that you, my supposedly namuh interlocutor, are not yourself one of these non-htraens? For all you know, you could be one also. So think carefully before you deny moral entitlements to the non-htraens, you may also be denying them to yourself.

On the basis of this, I think we might have reasonable confidence in the namuhs abandoning the claim that species membership is the criterion of moral considerability. What reason do we have for this confidence? Simply that in a similar situation – a situation in which what was thought to be the class of human beings is in fact made up of two genetically distinct species – we would do the same. Our reasons for this would, of course, be partly self-interested ones. If you do not know whether you will turn out to be a member of the deviant group, it would not be wise to endorse a rule which, by stripping members of that group of their entitlements, would also potentially result in the loss of your moral entitlements. There is, however, also a deeper intuition underlying this claim; an intuition that is not grounded in self-interest. If a person were to be discovered to be a member of a genetically deviant group, and, in virtue of this, not a member of the species *homo sapiens*, there is still a clear sense in which they are the same person they always were. What has been revised is one of the biological categories to which they belong, but this revision has not changed their identity as the particular person they are. Suppose it happened to you, for example.

You spend all of your life regarding yourself as an ordinary member of the species *homo sapiens*. Then one morning you wake up to find that, in your sleep, scientists have been running genetic tests on you. They have discovered that you are in fact a genetic anomaly: although you possess all the typical phenotypic properties of a human being, genetically you are not human at all. And since species membership is a genetic concept, this means that you are, in fact, not a member of the human race. However you might have conceived of yourself in the past, you are not a member of the species *homo sapiens*. There is a clear sense, I think, in which, despite the revision of one of the biological kinds into which you can be placed, you are still the same person you always were. After all, from your perspective nothing much has changed. You still remember things that happened to you in the past, you still have the same interests, thoughts, feelings, emotions, character traits, behavioural dispositions, and so on. In fact, if the scientists had not informed you of their discovery, you would, in all likelihood, have spent the rest of your life thinking of yourself as a human being. Intuitively, then, you are still the same person you always were. And, crucially, whatever aspects or features of you that go into making you deserving of consideration and respect, these aspects or features have not suddenly vanished with the mere discovery that assessment of your relation to a particular biological kind has been revised. As subjectively earth-shattering as it would no doubt be, the discovery is not sufficiently conceptually earth-shattering to change either your identity as the particular person you are or the moral entitlements you possess.

The conclusion we must draw from this is that species membership is not a morally relevant property. The fact that two individuals might be members of different species is not, by itself, sufficient justification for treating them differently. In particular, a difference in species membership is no justification for treating one of the individuals with consideration and respect while withholding such treatment from the other.

The above example is, to say the least, rather far-fetched (as, of course, is the whole Independence Day scenario). But the outlandish character of the example is not a significant drawback. What we are examining by way of such examples is the *concept* of moral considerability. That is, we are examining how we *think* about moral entitlements. And one way of facilitating this process is by considering various counterfactual examples and, often, the more outlandish the example, the more helpful it can be in this regard. We could also use the same process with regard to more mundane and well-understood concepts. We might, for example, examine the concept of a *bachelor* (to take a tried and tested philosophical example) by imagining various features that bachelors might have. Then, when we find that we cannot, without contradicting ourselves, talk about married bachelors, this tells us something important about the concept of a bachelor: the concept of a bachelor logically excludes the concept of being married. The reason that we do not do this with mundane and well-understood concepts is precisely because they are mundane and well understood. There is no point in clarifying these sorts of concepts; we already understand them perfectly well. However, part of what we are trying to do in the case of moral inquiry is clarify the nature of the moral concepts we employ. And in this case, the concept we are trying to clarify is that of moral considerability; of what it means to be morally entitled to consideration and respect. In this light, what the above thought experiments seem to tell us is that the concept of moral considerability is not closely tied to the concept of species membership. In certain circumstances at least, we would be quite happy to allow that the umbrella of moral considerability extends beyond the boundary of our species. Species membership cannot be regarded as the criterion of moral considerability because it is at least possible that there could be morally considerable individuals who are not members of the species *homo sapiens*. We can hope that the namuhs will follow us in this assessment, and accept the analogical extension to their own case.

Phenotypic properties

A possible namuh response to the problem of genetically distinct, but phenotypically identical, groups is to revise their criterion of moral considerability, now giving it a basis not in genetic properties but in *phenotypic* ones. On this suggestion, it is the physical appearance associated with being namuh that is decisive in determining who possesses moral entitlements in general, and the entitlement to consideration and respect in particular. Thus, it is the phenotypic properties associated with being namuh that constitute the criterion of moral considerability. Therefore, it is the different phenotypic properties possessed by namuhs and humans that provides the crucial morally relevant difference between members of the two species. And this is why all namuhs must be treated with equal consideration and respect while all humans can be sent off to the factory farm.

While this suggestion might be put forward by certain sections of the namuh population, it must be said that the more reflective namuhs would not be really happy with it. What troubles them is the possibility that there could be born a namuh who is so hideously deformed (from the namuh point of view) that he shares very few of the typical phenotypic properties of normal healthy namuhs. Nonetheless, despite his physical disfigurement, his mental powers are the same as those of a normal, healthy, namuh. This namuh is quite happy with this his life, enjoys the company of other namuhs (and as all namuhs admit, is a most agreeable host), and likes to engage in abstruse scientific and philosophical speculation. Surely, the more reflective namuhs think, it would be wrong to regard this unfortunate namuh as lacking any moral entitlements, hence as not being entitled to equal consideration and respect. To regard the namuh in this way, simply on the basis of his physical deformities, would be a particularly nasty form of chauvinism. And namuhs do not like to think of themselves as chauvinists. But, if the deformed namuh is morally considerable, despite his failure to possess a large proportion of the phenotypic properties of the average namuh,

then phenotypic properties cannot provide the criterion of moral considerability. Phenotypic properties, that is, cannot be crucial in determining who, or what, possesses moral entitlements.

Intelligence

The more philosophically sophisticated namuhs will be unruffled by the failure to identify either a genetic or phenotypic criterion of moral considerability. They never expected there to be such a criterion. It is more realistic, they claim, to suppose that the morally relevant difference between namuhs and humans lies not in physical differences but in psychological ones. In particular, the vast gulf in intelligence between namuhs and humans is the morally relevant difference which justifies them treating all namuhs with equal consideration and respect, and sending all humans off for slaughter. Namuhs are, while humans are not, morally considerable because the latter fall below the necessary threshold of intelligence.

There is, however, a serious problem with this line of argument also. If this is the namuh case against us, then we can employ what is known as the *argument from marginal cases* against them. That is, we should point out that, while it may be true that *most* namuhs are more intelligent than *most* humans, there is a class of namuhs for which this is not so. Firstly, some namuhs, for example, are born with severe brain damage resulting in severe retardation of their intellectual powers. Secondly, the intelligence levels of namuh infants are not noticeably different from those of adult humans. Finally, many ageing namuhs, through a variety of causes, suffer from a progressive deterioration in brain structure and function, and the intelligence of these is certainly no greater, and in many advanced cases less, than that of adult humans. Therefore, if the namuhs want to claim that humans, because of their inferior intelligence, lack moral entitlements, including the entitlement to respectful treatment, then it seems that, if they are to be consistent, they must also claim that these classes of namuhs lack

such entitlements. Happily for us, most namuhs are not willing to endorse this latter claim, and they accept, therefore, that they must reject intelligence as the criterion of moral entitlement. Since there is no difference in intelligence between these classes of namuhs and adult humans, intelligence cannot be a morally relevant difference between the two. Therefore, if the namuhs want to regard the unfortunate class of namuhs as morally considerable, while denying this status to adult humans, they cannot be relying on a criterion of intelligence in making this judgement. As far as any criterion based on intelligence is concerned, if the former are morally considerable then the latter must be also.

This argument from marginal cases, in fact, gives us an extremely powerful negotiating instrument. The reason is that the argument is completely neutral with respect to the property advanced as the criterion of moral entitlement. If, for example, the namuhs claimed that the morally relevant difference between them and us was their ability to appreciate *cisum*, an auditory art form of which humans had no comprehension, then all we would have to do in order to apply the argument from marginal cases is show that there are at least some namuhs who are unable to appreciate *cisum*, or whose ability to appreciate *cisum* is no greater than that of humans. If we can do this, then the namuhs face a choice: either they deny any moral entitlements to the relevant class of namuhs, or they abandon the capacity to appreciate *cisum* as the criterion of moral entitlement. Thus, the argument from marginal cases is completely neutral with respect to any proposed criterion of moral entitlement, and can, therefore, be applied independently of any particular criterion. The argument has the following general form of a dilemma:

- (1) X is proposed, by group G, as the criterion of moral entitlement.
- (2) There are certain members of G which do not possess X.
- (3) Therefore, either (a) those members of G possess no moral entitlements, or (b) X must be abandoned as the criterion of moral entitlement.

The argument is applicable no matter what X is, and no matter what the group G that is proposing X as the criterion of moral entitlement. Much of the power of the argument lies in its generality.

We humans, of course, will be hoping that the namuhs adopt option 3(b). And, happily for us, I think many of the namuhs would be willing to take this option. However, this may not be true for all of them. Some namuhs, in fact, may be willing to adopt option 3(a). According to proponents of this option, the fact that some namuh fail to measure up to the levels of intelligence required for possession of moral entitlements does not entail that the intelligence criterion should be abandoned. All it means is: so much the worse for those namuh. These unfortunate namuh are not, in fact, genuine possessors of moral entitlements. They are not, in reality, morally entitled to anything. To make his position more palatable, the defender of this view might employ a distinction between being a *direct* and being an *indirect* bearer of moral entitlements. The namuh who measure up to whatever level of intelligence is demanded by the criterion are direct bearers of moral entitlements. The namuh who fail to measure up lack such entitlements, at least directly. However, it is also true that the namuh who are direct bearers of rights may, in various ways, be attached to those who are not. In various ways, and to various degrees, they sympathize with their brothers, sisters, sons and daughters, mothers and fathers who fail to meet the requirements of the criterion. Moreover, in various ways, and to various degrees, they would be upset should harm befall these unfortunate namuh. Furthermore, as it has often been noted (most famously by their namuh philosopher, Tnak), the namuh who is cruel, callous, or indifferent in his dealings with those less fortunate namuh who are not direct bearers of rights tends also to be more cruel, callous, or indifferent in his dealings with those namuh who are the bearers of direct rights. If we sanction harm being done to the less fortunate namuh, we immediately put ourselves on a slippery slope, the inevitable conclusion of which is that harm will be done to those more fortunate namuh who are direct possessors of moral entitlements. Therefore,

the namuh defender of option 3(a) could argue that although those namuh who fail to meet the requirements of the criterion thereby fail to be direct bearers of moral entitlements, they do, nonetheless, possess such entitlements in a derivative sense. Since harm done to them can result in harm being done to those namuh who are bearers of direct rights, harm should not be done to those namuh who lack moral entitlements. This is not because they are entitled to not being unnecessarily harmed, but because of the connections they have to namuh who are entitled to not being unnecessarily harmed. Thus, the namuh who fail to measure up to the level of intelligence demanded by the criterion, it could be argued, possess entitlements in a derivative sense. They are indirect possessors of entitlements.

Even with this qualification in place, however, option 3(a), and the intelligence criterion upon which it is based, faces formidable difficulties. In particular, it is difficult to see how intelligence can be the crucial feature determining which entities are, and which entities are not, possessors of moral entitlements, direct or otherwise. Part of the worry here is that there are fundamental difficulties in identifying any non-arbitrary level of intelligence as constituting the criterion of moral entitlement. Suppose, for example, the namuh were to be attacked by yet another species who are significantly more intelligent than them. Whereas the average namuh IQ is, let us suppose, about 900, the average IQ of this invading species is more than double this. Consequently this new species regards the namuh in much the same way as they regard us, and as we regard other higher mammals. The new species, we will suppose, are also *intelligencists*: they regard a certain level of intelligence as constituting the criterion of moral considerability, and they identify this level as significantly higher than that of the most intelligent namuh – 1500 on the human scale, let us suppose. But what justification could there be for identifying this as the crucial threshold level? Why 1500? Why not, say, 700? Or 100? Or 10? What does any one of these numbers have to recommend it over any of the others?

This problem, it should be recognized, is not an epistemological one. That is, it is not simply a problem about the limits of our knowledge. It is not as if there is a certain numerical level of intelligence which provides the morally relevant threshold but that we just can't work out what it is. The problem is that any number we pick, any level of intelligence we identify, is equally arbitrary. There is nothing, no fact of the matter, that could recommend any one number over any other. And, since any particular level of intelligence we happen to identify, not just seems, but actually is, equally arbitrary, this shows that intelligence cannot be the crucial factor determining who possesses moral entitlements. Therefore, option 3(a) should be rejected.

4 Thou art the man

The biblical passage quoted at the beginning of the chapter pertains to the behaviour of David, the king of Israel. David apparently lusted after Bathsheba who, unfortunately from David's point of view, was married to Uriah. Not to let a little thing like this thwart him, he ordered Uriah off to fight in the front line of his army thus, effectively, ensuring his death. The point of Nathan's anecdote, of course, is to get David to see his behaviour from a new perspective, one unclouded by his libidinal promptings. Thus, David comes to see his behaviour not as a clever piece of manoeuvring in the game of love, but as a cruel and vicious act against an essentially defenceless opponent.

The Independence Day scenario is, of course, a thinly veiled parable for our treatment of animals (subtlety, I'm afraid, was never my strong point). For, with respect to our treatment of non-human animals, we are the *namuh*, the animals are us. And, in the argument we might have used against the *namuh*, we have the essential argument for animal liberation, where, for the present, we can understand this simply as the attempt to show that non-human animals possess substantially more moral entitlements than we

humans typically accord them. The argument has the following general form:

- P1. Human beings possess a substantial set of moral entitlements including, fundamentally, the entitlement to equal consideration and respect.
- P2. There are no morally relevant differences between humans and non-human animals.
- C. Therefore, non-human animals also possess a substantial set of moral entitlements including, fundamentally, the entitlement to equal consideration and respect.

Defenders of the idea of animal liberation often point out the conceptual connections between this idea and the claims of certain oppressed groups, arguing that just as the treatment of one group by another can be, and often is, racist or sexist in character, so too the treatment of non-humans by humans is typically *speciesist* in nature. Remarkably, defenders of animal liberation are sometimes attacked for this claim, as if it somehow demeaned the fight against racism or sexism to have non-human animals mentioned in the same breath. However, the comparison is perfectly legitimate. This can be seen simply from the fact that the argument described above is simply one instance of a more general argument form:

- P1. Individual members of group X possess a substantial set of moral entitlements including, fundamentally, the entitlement to equal consideration and respect.
- P2. There are no morally relevant differences between individual members of group X and individual members of group Y.
- C. Therefore, individual members of group Y also possess a substantial set of moral entitlements including, fundamentally, the entitlement to equal consideration and respect.

It is because the fundamental argument for animal liberation is an instance of this more general form that we could make use

of essentially the same argument against the namuh in a bid to achieve human liberation. And it is because the argument for animal liberation is an instance of this more general form that the comparison with arguments against racism and sexism is both legitimate and logically compelling. If there is a difference between the cases against racism and sexism on the one hand, and the case against speciesism on the other, it is not in virtue of the logical form of the arguments these cases instantiate. The logical form is, in each case, identical.

There may, of course, be other reasons for thinking that the case against speciesism is disanalogous to the cases against racism and sexism. If one could show, for example, that there were morally relevant differences between humans and non-humans, while there were not between male humans and female humans, or between white humans and non-white humans, then one would have justification for one's claim that the cases are disanalogous. In any event, the crucial premise in the above argument for animal liberation is P2: the claim that there are no morally relevant differences between human and non-human animals which could justify the claim that the former are morally entitled to be treated with consideration and respect while the latter are not.

With respect to P2, much of the arguments of animal liberationists have been essentially, and necessarily, defensive in character. That is, they consist in demolishing suggestions of in what the morally relevant difference might consist. The most common suggestions of anti-liberationists, in this regard, mirror those of our imagined namuh interlocutors. These suggestions, as we have seen, can be objected to on the grounds that they are either morally arbitrary (genotype, phenotype, intelligence) or fall victim to the argument from marginal cases (intelligence, and most other qualities possessed by typical humans). And, it remains true that opponents of animal liberation have failed to put forward any satisfactory suggestion for in what the morally relevant difference between humans and non-humans might consist. Therefore, I think, the burden of argument is clearly with those who want

to maintain that there is a morally relevant difference between human and non-human species.

In this book, despite the indisputable success for the animal liberationist of the negative tactic, I want to explore a more positive approach. This approach consists in examining the other crucial, but obviously neglected, concept involved in the animal liberation argument. This is the idea of *treating individuals with equal consideration and respect*. If we want to know what is involved in treating non-humans with consideration and respect equal to that accorded humans, we shall first have to know what is involved in treating humans with equal consideration and respect. Moreover, I shall argue that once we understand what is involved in treating humans with equal consideration and respect, and once we understand the basis of the requirement to do so, we shall automatically see that (at least some sorts of) non-humans must be treated with consideration and respect equal to that which we accord ourselves. That is, proper understanding of the concept of equal consideration itself reveals that there are no morally relevant differences between human and non-human animals. This is the central plank of the contractarian defence of animals I shall develop in Chapter 6.

5 Who speaks for wolf?

Since the chapter opened with a parable, it seems appropriate to close with one. I make no claims for the historical accuracy of this story, indeed, it strikes me as extremely implausible. But, historical veracity, of course, is not its point. The parable runs as follows. For their council meetings, Native Americans of the Iroquois nation were reported to have had a person whose function was to represent not the interests of any particular tribe, but those of the non-human inhabitants of the plains; the four legged creatures, the winged creatures, the crawling creatures. This person would be invited to speak by the question: who speaks for wolf? This book is, in effect, a piece of philosophical advocacy, an attempt to speak for wolf. But

with such an attempt, sentiment is, at least in philosophical contexts, inappropriate. And this somewhat misty-eyed invocation of the wisdom and decency of aboriginal Americans is the last thing even approaching sentimentality that will be found in this book. Any piece of philosophical advocacy must stand or fall on nothing other than the validity and soundness of its arguments. And if the arguments of this book are correct, our treatment of many non-human creatures – four-legged, winged, and crawling – is of a piece with, and morally no more defensible than, the namuhs' imagined treatment of us. And, in the absence of morally relevant differences, to speak for humans is, logically, to speak for wolf.

3

Utilitarianism and Animals: Peter Singer's Case for Animal Liberation

Peter Singer can, with justification, be regarded as the founding father of the contemporary animal liberation movement. The increased public awareness of what exactly transpires in our treatment of non-humans – in factory farming, medical research, product testing, and so on – is, to a significant extent, due to the wide circulation of his work. Consequently, anyone who cares about the welfare of non-human animals must acknowledge an enormous debt to Singer. However, it is important to distinguish the beneficial impact Singer's work has had on public awareness from the philosophical arguments he uses to defend the moral claims of non-humans. The two are logically independent of each other. And this chapter is concerned purely with the philosophical arguments.

In the opening chapter of *Animal Liberation*, Singer presents a powerful argument for the claim that justice requires equal consideration of the interests of non-human as well as human animals.¹ And Singer also places the idea of equal consideration at the centre of the conceptual stage. The notion of moral equality, and the requirement of equal consideration that stems from it, Singer argues, is not a *description* of how the world is. Whether we like it or not, we humans come in different shapes and sizes, with different intellectual abilities, different moral capacities, different capacities to experience pleasure and pain, and so on. Thus, if the demand for equal consideration were based on the actual equality of human beings, it would be a manifestly unrealistic demand.

In fact, however, the concept of equal consideration derives not from any actual equality between human beings. Its function is not to describe human beings, but *prescribe* how we should treat them. Then, given this understanding of the nature of the requirement of equal consideration, Singer goes on to argue that we are also committed to applying the requirement in our treatment of non-humans.

As I have hopefully made clear in the preceding chapters, I believe that this is precisely the right strategy. That is, the case for the moral claims of non-humans turns decisively around the concept of equal consideration. The central flaw in Singer's case, I shall argue, stems not from its *adoption* of this strategy, but from its *implementation*. Singer, I shall argue, operates with an inadequate understanding of the concept of equal consideration. The reason for this, as is made clear in some of his other writings, is that Singer is a *utilitarian*, and interprets the concept of equal consideration in terms of the central principles of utilitarian moral theory.² The principal aim of this chapter is to examine utilitarianism, in its various forms, with particular reference to the idea of equal consideration which it supports. I shall argue that the concept of equal consideration underwritten by utilitarian moral theory is crucially deficient and that this, effectively, undermines Singer's case for animal liberation.

1 Utilitarianism I: definition of utility

Utilitarianism, in all its forms, is essentially made up of two separate components:

1. A definition of human welfare, or *utility*.
2. A requirement to *maximize* utility.

This section deals with utilitarian ideas of human welfare. The next section deals with why utilitarians think we should maximize it.

There are two clearly distinct conceptions of human welfare evident in utilitarian writings. The first identifies such welfare with *pleasure*, or, more generally, with *happiness*. The second identifies it with the satisfaction of preferences. The former is known as *hedonistic utilitarianism*; the latter as *preference utilitarianism*.

Hedonistic utilitarianism

Hedonistic utilitarianism has, historically, been perhaps the most influential in the utilitarian tradition. In its more restricted versions, this view claims that human welfare principally consists in the sensation or experience of pleasure. Pleasure is the primary human good because it is the one good which is an end in itself, to which all other goods are merely means. Bentham, one of the founders of utilitarianism, notoriously said that 'pushpin is as good as poetry' if it gives the same intensity and duration of pleasure. Poetry is better than pushpin only if it gives people more pleasure.

The problems with hedonistic utilitarianism are well known. Jack Smart, for example, asks us to imagine a pleasure machine, a device which injects drugs into us, thus creating in us the most pleasurable mental states imaginable.³ It seems that if pleasure were our greatest good, then logically we should all want to be hooked up to this machine for ever, living a life of nothing but intense pleasure (suppose, to make the choice stark, once you are hooked up, you cannot be unhooked). The problem is, however, that, in the eyes of most philosophers at least, volunteers for this procedure would be few and far between. In the eyes of many critics of utilitarianism, such a life would not be a life most worth living; on the contrary it would be a rather sad waste of a life. And if this is true, pleasure cannot be our ultimate value.⁴

For these sorts of reasons, many hedonistic utilitarians reject the identification of welfare with pleasure. Pleasure, it is argued, is simply far too restricted in content to be the principal human good. For many of us at least, it is argued, the things worth doing and having in life are not all reducible to one mental state like pleasure.

Not all valuable experiences need be pleasurable. On the contrary, many different kinds of experiences are valuable, and the best sort of life is one which promotes the entire range of valuable mental states. And utilitarianism, therefore, should be concerned with all valuable experiences, whatever form they take. If we denote the set of all such experiences with the term 'happiness', then we can reformulate hedonistic utilitarianism as the claim that human welfare principally consists in happiness.

This extended notion of human welfare, however, does not allow the utilitarian to avoid Smart's objection. To see this, all we need do is imagine the machine suitably extended in its capacities to produce any desired mental state – not just intrinsically pleasurable ones.⁵ Thus, the machine might be able to produce the experience of love, the sense of accomplishment from achieving a difficult task, and so on. Even with this extension of its capacities, many critics of hedonistic utilitarianism contend that we would still be hard pressed to find any volunteers willing to be hooked up for life. What we want in life, they claim, is something more than simply the acquisition of any kind of mental state. We do not just want the sense of satisfaction engendered by accomplishing a difficult task, we want to accomplish the task itself. We do not simply want the experience of writing a respected philosophical tract, or a Booker prize-winning novel; we want to actually write these things. What this seems to show, these critics contend, is that while we do find certain sorts of experiences valuable, we also find other things, over and above experiences, to be valuable. Moreover, the value of these sorts of things is of a sort that is not replaceable by the value of the experiences that might accompany them. This is why not many of us, it is claimed, would volunteer to be hooked up for life to the suitably extended experience machine.

Preference utilitarianism

According to the most straightforward version of preference utilitarianism, human welfare consists in the satisfaction of human

preferences, whatever these might happen to be. A person might want the experience of accomplishing a difficult task, or writing a Booker prize-winning novel; and these are preferences that could be satisfied in Smart's machine. However, a person might also actually want to accomplish the task or write the novel; and this is a preference that cannot be satisfied in the machine. According to the most basic form of preference utilitarianism, then, the principal welfare of a human being consists in the satisfaction of his or her preferences, whatever those preferences might be.

The problem with this view, as stated, is that, intuitively, our preferences do not always contribute to our welfare. I might have a preference to drink vast quantities of alcohol every night, but this would probably not contribute to my general welfare. Our welfare and our preferences do not necessarily coincide. This, by itself, does not show that there is anything more to our welfare than our preferences. After all, the reason why my current preference for large amounts of alcohol is not in my long-term welfare is that it conflicts with other preferences of mine – to stay healthy, live a reasonably long life, hold on to my job, and retain a viable bank balance, for example. But what this does show is that our preferences can, often without us realizing it, conflict. And, given this is so, not *all* of our preferences can coincide with our welfare.

A more sophisticated version of preference utilitarianism attempts to accommodate the problem of mistaken preferences by defining welfare as the satisfaction of *informed* or *rational* preferences. Human welfare, on this view, consists in the satisfaction of informed human preferences; that is, satisfaction of those preferences which are based on full information and correct judgments, while rejecting those that are mistaken or irrational. This position does seem far more plausible. However, it also seems extremely vague. It places very few constraints on what counts as utility. Pleasure at least had the merit of being reasonably easy to identify, and its presence easy to discern in others. We also have a reasonably good idea of how to promote pleasure. Once, however, we view utility in terms of informed or rational preferences, these features

disappear. There are many different kinds of informed preferences, and no obvious way of aggregating them. How do we know whether to promote poetry or pushpin if there is no single overarching value, like pleasure, by which to measure them? And, more generally, how do we know what preferences people would have if they were informed and rational?

There are, of course, genuine difficulties here. However, it would be unfair to use them to criticize utilitarianism, since the same difficulties emerge in connection with almost any moral theory. Ultimately, every moral theory has to confront the difficult issue of the nature of human welfare. The problems here, therefore, are not peculiar to utilitarianism. Indeed, nothing prevents the utilitarian from adopting whatever account of welfare her critics favour. If there is a damaging objection to utilitarianism, therefore, it will have to be found in the second part of the theory – that is, in the instruction to maximize utility, whichever definition of utility we finally adopt.

2 Utilitarianism II: maximizing utility

The second component of utilitarianism is the instruction to maximize human welfare or utility, however this latter notion is defined. According to utilitarianism, the morally right action, in any given situation, is the one that maximizes utility – for example, produces the greatest amount of happiness, or satisfies as many informed preferences as possible. In any such maximization, some people's preferences will necessarily go unsatisfied; the satisfaction of the preferences of some people is incompatible with the satisfaction of those of others. Therefore, if a person's preferences conflict with what maximizes utility overall, then that person's preferences will, in this situation, be overridden. However, since the number of preferences satisfied necessarily outnumber the number of preferences frustrated, there is no reason, utilitarians claim, why the preferences of the losers should take precedence over the more

numerous, or more intense, preferences of the winners. Notice also, that to override a person's preferences in a given situation is not, according to the utilitarian, the same as to ignore those preferences. The preferences have been considered and added into the preference calculus which determines the best course of action; that is, the course of action which results in the satisfaction of the greatest number of preferences. It is just that, once the results of the calculation have been determined, some preferences will have to be sacrificed in order to bring about the greatest number of satisfied preferences. The overriding of some preferences, then, is perfectly compatible with the equal consideration of all preferences.

There are two distinct arguments for maximizing utility that can be identified in utilitarian writings. These arguments are quite distinct, indeed even incompatible, and they generate two very different interpretations of utilitarianism. The difference turns on the conceptual role of the notion of justice in each interpretation.

Teleological utilitarianism: individuals as receptacles

John Rawls defines a teleological moral theory as one which (i) provides an independent definition of the good for humans, and (ii) defines justice in terms of the maximization of the good.⁶ This, according to Rawls, is precisely what utilitarianism does; and for this reason, we can refer to this interpretation of utilitarianism as the *teleological* interpretation.

According to the teleological interpretation, our primary moral duty is to maximize utility, however this is defined. That is, our primary duty is not to treat people as equals, but to bring about valuable situations or states of affairs; and the more utility possessed by a state of affairs, the more valuable that state of affairs is. On this view, people don't have value intrinsically, but they are, or can be, the *bearers* of what has value. These intrinsically valuable things will be states such as happiness, or preference-satisfaction, depending on one's view of what utility is. And the prime moral directive of utilitarianism is to maximize the number or amount of these

intrinsically valuable things. How we maximize these things, and, in particular, in whom we maximize them, is of no direct moral concern. Thus, on this view, people can be viewed as *receptacles* of what has value, and not things which are themselves intrinsically valuable. In this connection, Regan introduces a rather striking and helpful analogy: we can compare people to cups that are capable of containing either sweet-tasting liquid (utility) or bitter-tasting liquid (disutility).⁷ The primary goal of moral action is to produce as much sweet-tasting liquid, and as little bitter-tasting liquid, as possible. It doesn't matter, at least not directly, into which cups we pour the sweet-tasting liquid, as long as we introduce as much sweet-tasting liquid into the world as possible. Thus, if, for example, it should prove necessary to fill some cups entirely with bitter-tasting liquid so as to produce the maximum amount of sweet-tasting liquid elsewhere, then this is what we should do. In fact, if the maximization of the quantity of sweet-tasting liquid required filling some cups entirely with bitter-tasting liquid, then this is what justice requires we do. Similarly, if maximization of utility requires sacrificing the welfare of certain individuals, then justice requires that we sacrifice these individuals. This is so because justice, according to the teleological interpretation of utilitarianism, consists in the maximization of utility. Individuals who are sacrificed in this way, then, cannot, according to the teleological interpretation, complain that they are being treated unjustly.

The teleological interpretation is actually found in very few utilitarian writings – G. E. Moore being its only notable advocate.⁸ Indeed, the teleological interpretation seems to be one primarily found in the writings of critics of utilitarianism. As mentioned above, Rawls, for example, sees utilitarianism as fundamentally a teleological theory in this sense.⁹ And this is the basis of his charge that utilitarianism essentially collapses the distinction between different people. Tom Regan, another vociferous critic of utilitarianism, also sees it as primarily a theory of the teleological sort.¹⁰ This provides the basis of his claim that utilitarianism treats people as merely receptacles of what has value, and not as inherently valuable

in themselves. The teleological interpretation, then, seems largely to be an interpretation foisted on utilitarianism by its critics.

There are, in fact, good reasons why utilitarians should not adopt the teleological interpretation. One of the things a moral theory attempts to do is not only to spell out a set of duties that a moral agent can reasonably be thought to possess, but also to identify those individuals towards whom the agent has these duties. The teleological interpretation of utilitarianism is committed to the claim that we have a duty to maximize valuable states of affairs, but leaves it wholly mysterious as to whom we have this duty. On the teleological interpretation of utilitarianism, there seems to be no identifiable individual or group of individuals who can plausibly be regarded as the beneficiary of such a duty. It is implausible to claim that we have this duty to the maximally valuable state of affairs itself, for it is not clear how states of affairs can have moral claims. The most natural response is to claim that we have the duty to whichever individuals would benefit from the maximization of utility. But then it seems our primary moral duty is to individuals, and that we have only a derivative duty to produce maximally valuable states of affairs. And in this case, the teleological interpretation of utilitarianism collapses into another, quite distinct, version of that theory. It is to this second interpretation that we now turn.

Egalitarian utilitarianism: individuals as counting equally

The teleological interpretation of utilitarianism begins by defining human welfare – happiness, preference-satisfaction, and so on – and then defines justice as the maximization of this welfare. The concept of justice is derivative upon the concept of welfare. What we can call the *egalitarian* interpretation of utilitarianism, on the other hand, treats the concept of justice as basic, and the requirement to maximize utility is then derived from this. Thus, in the writings of many utilitarians, we find the following sort of argument. Each person should be regarded as a moral equal, and we must, therefore, treat everyone with equal consideration and respect. To

this end, the preferences of each person should be regarded as having equal weight, irrespective of the content of their preferences, and regardless of the specific talents, capacities, endowments, and physical and economic circumstances of the person. Giving these preferences equal weight is what is required to treat people as equals, to treat them with equal consideration and respect.

In other words, on the egalitarian interpretation, we can regard the content of utilitarianism as being factored into three components. The first is a formal principle of justice:

- U1. Each person should be treated with equal consideration and respect.

This principle, as we have seen, is generally regarded as constitutive of, or essential to, the moral point of view as such, and is in no way peculiar to utilitarianism. What is peculiar to utilitarianism, however, is its interpretation of this formal principle. This interpretation consists of two principles. The first is:

- U2. The interests of each and every person should be given equal weight in moral deliberations.

The notion of an interest, here, functions simply as a place-holder for the slightly more concrete concepts of pleasure, happiness, preference-satisfaction, and the like. U2 is an interpretation of U1. We are to treat each person with equal consideration by allowing their interests (e.g., their preferences) to count equally. The final principle is an interpretation of U2.

- U3. The maximum possible number of interests should be satisfied.

U3 is an interpretation of U2. According to U3, the best way to make each person's interests count equally is to ensure maximal satisfaction of interests.

The crucial difference between the teleological and the egalitarian interpretation of utilitarianism can be understood as follows. The teleological interpretation makes maximization of utility *constitutive* of justice. The egalitarian interpretation, on the other hand, begins with a prior conception of justice (all people should be treated with equal consideration and respect) and interprets this conception as the requirement that utility should be maximized. What is crucial to the egalitarian conception is that the requirement to maximize utility is derived entirely from the prior requirement to treat people with equal consideration. This sort of justification for maximizing utility can clearly be found in the writings of utilitarians such as Mill, Hare, Griffin, and Harsanyi.¹¹

The egalitarian interpretation is also clearly the interpretation adopted by Singer. In the opening chapter of *Animal Liberation*, as in other writings, Singer claims that his arguments for the moral claims of non-humans derive from what he refers to as 'the basic moral principle of equality'. This principle, he points out, is not based on any alleged factual equality between different people. Whether we like it or not, people come in different shapes and sizes, they come with different moral capacities, different intellectual abilities, different abilities to communicate effectively, and so on. In short, if the truth of the principle of equality depended on there being any factual equality between human beings, the principle would be certainly false. However, Singer argues, the principle does not function to *describe* any factual equality between humans. Its function is *prescription* not *description*. The principle prescribes how people should be treated, not describes how they are. Singer's utilitarianism, then, derives from his viewing the equal consideration of all interests as the best means of satisfying the basic moral principle of equality.

Since it is the egalitarian interpretation which is clearly adhered to by most defenders of utilitarianism, and since, as outlined above, there are significant problems with the teleological interpretation, in the sections to follow I shall focus on the egalitarian interpretation. Given this interpretation, the central problem with

utilitarianism can, I think, be stated as follows: *principles U2 and U3 provide a poor interpretation of the formal principle of justice U1*. The section to follow will develop arguments that attempt to show exactly why this is so. These arguments, however, can easily be adapted to apply to the teleological interpretation of utilitarianism. That is, although arguments will be presented as showing that utilitarianism provides a poor *interpretation* of the formal principle of justice, they can also be taken to show that utilitarian deliberations cannot, plausibly, be regarded as *constitutive* of the formal principle of justice.

3 Problems with utilitarianism

The principal motivation for utilitarianism, then, at least in the eyes of many of its advocates, is egalitarian; it is essentially motivated by a concern for treating all people with equal consideration. Indeed, as Hare says, if we believe that each person's welfare consists in the satisfaction of their informed preferences, then, in order to treat each person with equal consideration, what else can we do except give equal weight to their preferences, everyone counting for one, no one for more than one.¹² Thus, Hare sees utilitarianism as providing the only possible way of giving equal consideration to everyone. Underpinning the egalitarian conception of utilitarianism, therefore, is the inference from treating people with equal consideration to giving equal weight to each person's interests or preferences. The former, on the egalitarian conception of utilitarianism, requires the latter.

This inference, I shall argue, is questionable. I shall argue that to interpret the principle of equal consideration in terms of the maximization of satisfaction of interests is to badly misunderstand the principle of equal consideration. More precisely, U2 and U3 amount to what we can call an *aggregation requirement*. We are to treat people with equal consideration by giving equal moral weight to all their interests. And we give equal moral weight to all their

interests by satisfying the maximum possible number of them. And this means we have to aggregate interests and adopt whatever course of action is required to bring about the maximum number of satisfied interests. I shall argue that to understand the concept of equal consideration in terms of this sort of aggregation requirement is to badly misunderstand the concept. Utilitarianism, then, has misinterpreted the ideal of equal consideration. And, as a result, it allows some people to be treated as less than equals, as a means to other people's ends.

One central component of our intuitive conception of equal consideration is surely this: the moral entitlements an individual can be legitimately thought to possess do not depend on, and are in no way altered by, the attitudes that other individuals bear towards them. If I, for example, dislike Smith, and want him to be deprived of certain goods, resources, or opportunities to which he claims entitlement, then my attitudes, *by themselves*, do not entail that Smith should be deprived of these things. If Smith is genuinely entitled to these goods, resources, or opportunities, then my hostility towards him can do nothing to change this. Indeed, if everyone in the world harbours inimical feelings towards Smith, then this in no way alters his moral entitlements. Conversely, if everyone in the world positively adores Smith, this in no way increases his moral entitlements. We do, of course, sometimes restrict an individual's access to certain goods, resources, or opportunities, and often when we do so we also actively dislike the individual in question. Our treatment of certain criminals provides an obvious example. However, our active dislike of someone who has committed a particularly heinous crime is, in this sort of case, simply an *accompaniment* to the restrictions imposed on them, not a *justification*. The justification for restricting their goods, resources, and opportunities lies in the crime they have committed, not in the hostility this crime arouses in us. The claim that the moral entitlements possessed by an individual do not depend simply on the attitudes that other people bear towards him or her is essential to our intuitive understanding of the principle of equal consideration. One of the

problems with utilitarianism is that it is simply unable to accommodate this fact.

To see this, consider the following scenario. Suppose the vast majority of the populace of a society become increasingly desensitized to violence, perhaps due to the proliferation of violent films or whatever it is which is supposed to desensitize one to violence. And they decide that the usual offerings of contact sports – boxing, rugby, gridiron, hockey, and the like – simply don't hold the same fascination for them any more. What is needed is something far more violent. Therefore, in the quest for better weekend entertainment, network television decides to reinstitute the old Roman tradition of gladiatorial combat to the death. The viewing population who, you remember, has become severely desensitized to violence, is very excited about the idea. So too, therefore, are the advertisers. And so too, therefore, are the networks. The only problem is: finding the gladiators. The problem is solved when a young ambitious network executive hits upon the idea of using convicts on death row (or, in countries where there is no death penalty, those sentenced to life without the possibility of remission). The convicts are not given a choice; they are simply forced to fight.

Since the first edition of this book came out in 1998, this scenario is looking distinctly less outlandish than it did then. We may think of the so-called mixed martial arts of the UFC – that's the *ultimate fighting championship*, for the uninitiated – as a stepping-stone between traditional boxing and gladiatorial combat to the death. I suspect that one way in which the example is inaccurate is the supposition that we would have to work so hard to find contestants. The recent explosion of reality TV shows suggests that more than willing volunteers would be positively lining up to appear. But I'll stick with the original scenario in case anyone doubts the hold of reality TV fame over today's minds. In this scenario, we seem to have a situation where the preferences of a large number of people – the bloodthirsty populace of our imagined society – are set against the preferences of what is, in comparison, a vanishingly small number of people – the unwilling gladiators. If the

imagined society is the United States, for example, the population would number about 250 million, compared to, say, a few hundred convicts a year. In China, the disparity would be even greater. Even allowing for the fact that the preferences of the convicts are greater or more intense than those of the viewing population, it still seems that the preferences of this vast majority would outweigh those of the convicts. Therefore, it seems that utilitarianism would, in this sort of situation, be committed to the reintroduction of gladiatorial combat to the death as the morally right thing to do. But, intuitively at least, something here seems to be seriously unjust.

Of course, there could be complications in instituting such a programme, complications that the utilitarian might cite as tipping the preference calculus back against the gladiatorial system. Thus, for example, the gladiators might have friends or relatives who would be greatly saddened by their public slaughter. This sort of consideration is often referred to as a *side effect*. However, we can eliminate these sorts of complications by a slight articulation of the scenario. Suppose, for example, the names of the convicts were drawn by public lottery a week or so before their fight. Viewers then had a week to register a complaint, and the prisoner would go to the arena only if no one of the general populace objected. Presumably, gladiatorial candidates would be harder to find in these circumstances, but, in principle, there might well be enough friendless and family-less death row convicts to meet the network's requirements. The utilitarian, it seems, would be committed to claiming that, in this case, reintroduction of the gladiatorial system is not only morally legitimate but actually morally required.

The same sort of points can be made in relation to cases much closer to home. Some people, for example, are racists. And one form such racism might take is wanting a certain minority, or minorities, to have fewer goods, resources, and opportunities than one wants available to members of one's preferred racial group. One might, for example, want to exclude blacks from certain jobs, or certain educational opportunities, because one thinks they are not worthy of them. And if the minority in question were of sufficiently

small size, and if the racist members of society were sufficiently large in number relative to the minority group, and if the relevant beliefs of the racists were held with sufficient fervour, then taking proper account of the preferences of the racists might, on utilitarian grounds, justify repressing the minority group in the way preferred by the racists. Once again, whatever else we have here, we do not have a case of equal consideration. On the contrary, adopting the utilitarian calculus would entail that what treatment should be accorded the minority group is a function of what *other* people – in this case a group of racists – happen to think about them.

Furthermore, the situation would not be significantly changed if we replace the group of racists with a group of, say, benevolent despots who are much more favourably disposed to the minority group to the extent, for example, of even granting them preferred status with regard to employment and educational opportunities. The point remains the same. In this case, the treatment we accord the minority group is again a function of what other people happen to think about them; it's just that here the others happen to be a group who are more favourably disposed towards the minority group. The principle of equal consideration is incompatible with the idea that what a person, or a group of people, are rightfully owed is a matter of what other people happen to think about them. And this is true no matter what the specific content of those other people's attitudes; that is, no matter whether those attitudes are favourable or unfavourable.

What is going on in cases such as these is a clash between two principles which utilitarians have, arguably, conflated. On the one hand there is the principle of *equal consideration*: the principle that every person should be treated with equal consideration. On the other hand is the principle of the *aggregation of interests*: we are to aggregate the interests of all people and adopt whatever course of action maximally satisfies the aggregation. Utilitarians, as we have seen, interpret the former in terms of the latter. However, as the above examples make clear, the principles are not only non-equivalent, it is easy to imagine circumstances that reveal them to be incompatible.

At the root of the dissonance of the principle of equality with the principle of the aggregation of interests lies the fact that many interest are, as we might intuitively put it, morally *illegitimate* ones. My interest in watching unwilling gladiators fight to the death for my entertainment would be – if I had it – an illegitimate interest. So too would be my interest in ensuring that I and members of my own racial group get an advantageous distribution of goods, commodities, and freedoms. If we adopt the principle of the aggregation of interests, as do utilitarians, then this entails that what people are rightfully owed depends, in part, on what others decide they want them to have. But the desires of others in this regard need not be morally legitimate desires. This is why there is a clash between the principle of equal consideration and the principle of aggregation: the desires or interests that we end up aggregating can just as easily, and often do, turn out to be morally illegitimate desires or interests.

It may be objected, of course, that this above way of setting up the problem begs the question against the utilitarian. Utilitarians, after all, will deny that there is such a thing as ‘what one is owed’ or a ‘fair share’ independently of utilitarian calculations. That is, what counts as a morally legitimate or illegitimate interest is not something that can be determined prior to the results of utilitarian calculations; it can be determined only after these calculations have been made. Therefore, all desires or interests, no matter what their specific content, must be thrown into the utilitarian melting pot.

The problem with this objection, however, is that it relies heavily on what was earlier identified as a teleological interpretation of utilitarianism: the claim that the primary directive of utilitarianism is to maximize utility. However, it was argued that this interpretation is not the one adopted by most utilitarians. On the contrary, most utilitarians understand their theory as providing the best way to interpret a formal principle of justice, a principle motivated independently of considerations of utility maximization. And, if this is so, utilitarianism will involve a commitment to our intuitive understanding of the formal principle of justice. The real issue, therefore, is simply whether utilitarianism provides an

adequate interpretation of the principle of equal consideration. And, as the above cases hopefully make clear, it does not do so; at least it does not provide an adequate interpretation of our intuitive understanding of what this principle entails.

4 Utilitarianism and justice

It might be thought that the utilitarian can easily modify his position: we will exclude from utilitarian calculations all illegitimate desires or interests. Only legitimate desires or interests will be permitted to figure in utilitarian calculations. The problem, however, is that the utilitarian can give no substance to the distinction between legitimate and illegitimate desires or preferences. This distinction can only be drawn if we appeal to non-utilitarian principles of justice. And this, in effect, would entail the rejection of utilitarianism as a foundational moral theory.

The motivation for utilitarianism, it has been argued, is the principle of equal consideration. Utilitarians, however, interpret this principle in terms of the principle of the aggregation of interests. That is, to treat each individual with equal consideration requires giving equal weight to each of their interests or preferences. However, if we include morally illegitimate preferences among those to which we are to give equal weight, then the principle of the aggregation of interests is actually incompatible with the principle of equal consideration, at least as this is intuitively understood. By interpreting the principle of equal consideration in this way, therefore, utilitarianism, in effect, undermines its own motivation.

Nevertheless, it is also true that the interpretation of the principle of equal consideration in terms of the principle of the aggregation of interests is definitive of utilitarianism. For utilitarianism, in addition to being committed to the former, is also committed to the idea of a calculus, and therefore to things which can be weighed, aggregated, and measured against one another. It makes little sense to talk of entering individuals into the calculus in this sense.

So, utilitarianism needs things such as interests, or informed preferences, as suitable material for the calculus.

Therefore, not only is utilitarianism committed to the principle of equal consideration, it is also, it seems, committed to interpreting this principle in terms of the principle of aggregation of interests. Without this principle, it seems, we would lose much (or all) of what is distinctive about utilitarianism. The problem is, however, that, without the exclusion of illegitimate preferences, the principle of the aggregation of interests is actually incompatible with the principle of equal consideration, at least as this is intuitively understood. Therefore, on pain of violating our intuitive understanding of the concept of equal consideration, the utilitarian is committed to excluding illegitimate preferences from the principle of aggregation of interests. Not all preferences or interests are, in fact, to be given the same weight. Some are to be given no weight at all.

However, this creates a serious problem for the utilitarian. According to utilitarianism, which actions count as right, and which count as wrong, can emerge only *subsequent* to the calculus of interests or preferences. The possibility of moral evaluation, then, can emerge only as a result of the calculus. It therefore makes no sense, on utilitarian grounds, to exclude certain preferences on the grounds that they are illegitimate preferences *prior* to the calculus. For what counts as a legitimate or illegitimate preference can only be determined by the calculus itself. Therefore, the exclusion of illegitimate preferences, an exclusion which utilitarianism requires to safeguard our intuitive conception of equal consideration, cannot be justified on utilitarian grounds. It is an exclusion required by utilitarianism, but one that cannot in any way be motivated or justified by it.

This means that any exclusion of illegitimate preferences will have to be justified by a prior standard of legitimacy. And since utilitarianism is committed to this exclusion, utilitarianism is, therefore, also committed to a prior standard of legitimacy. But, then, utilitarianism cannot be thought of as providing the sole standard of morality. It must already incorporate, in a tacit manner,

prior standards of moral legitimacy. In other words, in order for utilitarianism to work, there would have to be prior, non-utilitarian, standards of morality. Utilitarianism, therefore, cannot provide the sole standard of morality.

John Rawls has argued that what is required, as a non-utilitarian standard of morality, is an adequate *theory of justice*: Rawls regards this as the fundamental difference between his account of justice and that of the utilitarians. For Rawls, it is a defining feature of our sense of justice that 'interests requiring the violation of justice have no value', and so the presence of illegitimate preferences 'cannot distort our claims upon one another'.¹³ An adequate theory of justice, for Rawls, limits the admissible conceptions of the good, and so those conceptions the pursuit of which violates the principles of justice are ruled out absolutely: the claims to pursue inadmissible conceptions have no weight at all. Because unfair or illegitimate preferences never enter into the preference calculus, people's claims are made secure from the unreasonable demands of others. Utilitarianism fails to exclude illegitimate preferences because utilitarianism is committed to interpreting the concept of equal consideration in terms of the aggregation of pre-existing preferences, whatever those preferences happen to be. And standards of equality can emerge only subsequent to the calculus. That is, what counts as equal consideration is, for the utilitarian, constituted by the results of the calculus itself. The problem, however, is that the notion of equality should enter into the very decision whether to take into account a particular preference; whether to enter it in the calculus in the first place. Thus, we need a prior standard of justice in order to even begin to engage in deliberation of consequences. Part of what it means to show equal consideration for others is taking into account what rightfully belongs to them. Hence illegitimate preferences must be excluded from the start, for they already reflect a failure to show equal consideration. Utilitarianism, however, can make no sense of this claim. Therefore, utilitarianism provides an inadequate interpretation of the principle of equal consideration.

5 Rule utilitarianism

The arguments of the previous section have tried to demonstrate the dissonance between the principle of equal consideration – the formal principle of justice – and the principle that the interest of all concerned should be aggregated. Some utilitarians, however, would argue that this dissonance is apparent rather than real. They accept that utilitarian reasoning can *appear* to have implications that are incompatible with the principle of equal consideration. But they claim that these implications can be avoided if we switch to a more sophisticated form of utilitarianism. According to this, it is not individual acts that are subject to the test of utility but, rather, the *rules* they embody. According to this view, since society is impossible without individuals adhering to rules, we should assess the consequences not simply of acting in a particular way on a particular occasion but of making it a rule that we act in this way. Thus, we should perform whatever act is required by the best rule; and the best rule is one consistent adherence to which will maximize utility. Thus, we are to adhere to utility maximizing rules, even when doing so results in our failing to perform utility maximizing acts. This view is known as *rule utilitarianism*. There are well-known problems with rule utilitarianism. Some have argued that rule utilitarianism ultimately collapses into act utilitarianism. To see this, suppose there is an action A1 which would, in a particular situation S, maximize utility. However, there is also a rule R consistent adherence to which also maximizes utility in the long run, and R, let us suppose, requires us to perform, in situation S, the distinct act A2. However, if this were the case, then it seems possible to replace R with another rule R* which states: 'perform act A2 unless in situation S; but when in situation S, perform A1'. This is a rule which seems to be more productive of utility than R since it inherits all the usual utility maximizing power of R and at the same time avoids the problem that R does not maximize utility in situation S. Thus, consistent adherence to R* would generate more utility than consistent adherence

to R. Therefore, it seems that the rule utilitarian is committed to adoption of R* rather than R. However, this procedure can be repeated for all those situations S2, S3, ..., Sn in which adherence to R results in less than maximal production of utility. And this will result in the further rules, R**, R***, and so on. And if this is so, there is a clear danger that our specification of the rules becomes so specific as to make rule utilitarianism indistinguishable from act utilitarianism.

Even if this problem can be surmounted, however, it is still very doubtful that rule utilitarianism can provide an adequate interpretation of our intuitive conception of equal consideration. The reason is that even rule utilitarianism makes the treatment a person is owed, as a matter of justice, contingent on the potentially illegitimate attitudes of other people. To see this, consider, for example, the rule utilitarian position with regard to illegitimate preferences such as racist ones. According to rule utilitarianism, the wrong done in discriminating against a minority group consists in the increased fear caused to others by having a rule allowing discrimination: fear based on the realization that 'I could be next!' But this claim is surely no closer to our intuitive conception of equality than is the corresponding claim of act utilitarians. The rule utilitarian claim entails that the treatment someone is owed, as a matter of justice, depends, at least in part, on the presence or absence of certain psychological attitudes in others. If discriminating against a minority group happens to result in increased fear in individuals outside that group, then it is morally wrong. If, on the other hand, such a rule did not result in this sort of increased fear, perhaps because the members of the majority group were too dim-witted to see the connection, then the rule should be regarded as morally legitimate. So, the rule utilitarian is also committed to the view that the treatment someone is owed, as a matter of justice, depends on the contingent presence or absence of certain psychological states in others. Rule utilitarianism, then, is no closer than act utilitarianism to our intuitive conception of equal consideration.

6 Utilitarianism and animals

The strength of utilitarianism, as a moral theory, is that it frequently, so to speak, gets the right answer. Its weakness is that it does so for the wrong reasons. That this is sometimes difficult to see is due to the fact that the sorts of situations necessary to highlight the problematic implications of utilitarianism are often, necessarily, outlandish. The example of the institution of gladiatorial combat provides an obvious example. In more realistic cases, racism for example, it is often difficult to see the implications of utilitarianism. It is difficult, for example, to imagine the desires of racists to restrict the opportunities of minorities possibly outweighing the preferences of those minorities not to be so restricted. The latter sorts of preferences would, by their very nature, presumably be far more intense than the former. Thus, in order to make the example at all plausible, we would have to imagine a vast disparity in size between the two groups. And so, once again, we veer towards the outlandish. Nevertheless, it remains true that, since utilitarianism, in both its act and rule forms, is committed to the claim that the treatment an individual is due as a matter of justice depends on the contingent psychological states of both that individual and others, utilitarianism, in both its forms, is incompatible with our intuitive understanding of the idea of equal consideration. Intuitively, to treat individuals with equal consideration requires treating them without regard to such contingent psychological features. Utilitarianism, then, often gets the intuitively right answer with respect to particular moral issues. It does so, however, for the wrong sorts of reasons; because, for example, of contingent features of the world such as the disparity in size between exploiting and exploited groups not being too great. Nowhere is this combination of right answer arrived at for the wrong reason more evident than in the utilitarian account of our moral commitments to non-humans. Let us consider Singer's utilitarian argument for vegetarianism.

The utilitarian case for vegetarianism is simple. If we are preference-utilitarians, for example, we will have to weigh up the

preferences satisfied and frustrated by a policy of continuing to eat meat against the preferences satisfied and frustrated by a policy of abandoning meat-eating. And since the preferences of those non-human animals involved in the animal husbandry process will have to be included, it may seem that utilitarianism licenses a straightforward and clear-cut result. Singer writes:

Since, as I have said, none of these practices (of raising animals intensively) caters for anything more than our pleasures of taste, our practice of rearing and killing other animals in order to eat them is a clear instance of the sacrifice of the most important interests of other beings in order to satisfy trivial interests of our own ... we must stop this practice, and each of us has a moral obligation to cease supporting this practice.¹⁴

As Singer sees it, the issue is simply one of weighing our relatively trivial preferences for gustatory satisfaction against the preferences of cattle, pigs, chickens, and the like, for a decent life free from undue suffering. And, seen in these terms, it is clear that utilitarianism would license the policy of abandoning meat-eating.

Tom Regan, however, feels that matters are not quite as clear-cut as this.¹⁵ After all, there are more human preferences involved than those of the merely gustatory sort. The animal industry is big business. It is uncertain exactly how many people are involved in it, both directly and indirectly, but certainly the number must run into the tens, and probably hundreds, of thousands. Firstly, there are those who actually raise and sell the animals. Then, there are the feed producers and retailers; cage manufacturers and designers; producers of growth stimulants and other chemicals; those who butcher, package and ship the produce. Then there are extension personnel and veterinarians whose lives revolve around the success or failure of the animal industry. Moreover, also to consider are all the members of the families who are the dependants of these employees or employers. The interests these people have in raising animals intensively go well beyond pleasures of taste and are far from trivial. These people have a stake in the animal industry as

rudimentary and important as having a job, feeding a family, and so on. When you add in these interests to the utilitarian calculations, things are perhaps not as clear-cut as Singer would have us believe.

Things become even murkier when we add in to the calculations the *side effects* which Singer, as a utilitarian, is obliged to take into account. Singer must take into account the preferences of *everyone* affected by the consequences of altering the animal industry, not just those who happen to be directly involved in it. The short- and long-term global economic consequences of a sudden or gradual transition to vegetarianism, must be investigated by any utilitarian. For example, it has been shown that the rate of inflation in countries such as the United States follows, quite closely, the price of beef. What would be the economic implications of a widespread abandoning of meat-eating? As the price of beef rose, as it almost certainly would in these circumstances, would there be a consequent rise in inflation, and, as a result, perhaps an increase in unemployment even for people not connected with the animal industry? These sorts of questions would have to be seriously addressed by the utilitarian. It is simply not enough to see the issue as a weighing up of the vital interests of animals over the trivial gustatory interests of human beings.

I think Regan's case is probably empirically implausible in this regard. That is, when you in fact weigh up the preferences of the humans directly and indirectly involved in the animal husbandry industry against the preferences of the vastly greater number of animals used in this industry, my suspicion is that the latter will significantly outweigh the former. Therefore, utilitarianism does entail that the practice of animal husbandry is morally wrong. And, as I shall argue in a later chapter, this conclusion is correct; the practice is seriously unjust. This is a case, then, where I think utilitarianism yields the right answer. The problem is that it does so for the wrong reasons.

To see this, imagine a few contingent changes in the circumstances underlying the practice of animal husbandry. By

imaginatively varying the circumstances, we can presumably imagine a case where the interests of humans outweigh those of animals. Indeed, given the aggregative nature of utilitarianism, it seems that there must necessarily be such a situation. All we have to do to arrive at such a situation is gradually reduce the number of animals involved in the industry, or gradually increase the interests humans have in the results of the process, or both, and, due to its aggregative nature utilitarianism entails that we must eventually reach a situation where the interests of humans outweigh those of animals.

To see this, consider the following scenario. Suppose animal protein had a different effect on humans than the one it in fact has. Suppose animal protein has no nutritional role in human growth or maintenance, but, instead, acts as a drug which induced in humans intense states of euphoria without the side effects associated with most euphoricants. Suppose also that humans could take very little animal protein at any given time, but that the effects lasted for weeks. Thus, far fewer animals were involved in the husbandry industry: numbering, say, only a few thousand worldwide at any given time. In this sort of situation, utilitarianism, it seems, would be committed to the idea that this level of animal husbandry is morally good. (If you don't think it would, just tinker around with the circumstances until you find a situation where you think it would.) Now, the central question, here, is not whether this would, in fact, actually be a morally good situation (although the theory to be developed in Chapter 6 entails that it is not). Rather, the central question, here, is what the fact that utilitarianism is committed to claiming that it is a good situation entails about the utilitarian concept of equal consideration. And, it is pretty clear, the utilitarian is committed to the following position: the treatment a human or animal is due is a function of the effects such treatment has on everyone affected by it. The animals who are involved in our imagined case are not being treated with equal consideration. The treatment they receive is a function not of any feature they possess in themselves, but of the effects of their treatment on others. They are

being treated simply as means. And to endorse this claim is to reject the principle of equal consideration.

Utilitarianism, I think we should conclude, can support no robust concept of equal consideration. Ultimately, utilitarianism is committed to the idea that the treatment an individual deserves is a function of the interests everyone – and not just the individual – has in such treatment. Thus, if a robust concept of equal consideration is to be found, it will have to be in other, non-utilitarian, moral theories. In the next chapter, we look at such a theory.

4

Tom Regan: Animal Rights as Natural Rights

In his seminal work *The Case for Animal Rights*, justifiably regarded as a classic of the animal liberation movement, Tom Regan presents a forceful and, in some ways, compelling account of why non-human animals should be regarded as making direct moral claims upon us. The reason, according to Regan, is that non-humans possess moral rights; and he presents an elegant and systematic theoretical underpinning for this claim. I think it is fair to regard Regan's case as proceeding from within the framework of natural rights approaches to morality. This is for three reasons. Firstly, Regan argues that many kinds of non-human animals possess moral rights in virtue of their *nature*; in virtue of the fact that they are, as he puts it, *subjects-of-a-life*. Secondly, his argument appeals quite centrally to the concept of *inherent value*, viewed as an objective moral property which attaches to certain things, and which does so irrespective of whether those things happen to be valued or not. Thus, Regan views at least some non-human animals as possessors of moral rights which are objective in the sense that they do not depend on whether they are recognized as rights. Thirdly, these rights are logically prior to any contractual arrangement, since they stem from the nature of the individuals and not from the agreements such individuals might enter into. In this sense at least, Regan is an inheritor of the conceptual framework embodied in natural rights doctrine.

One of the central claims of this book is also that at least some sorts of non-human animals possess moral rights. However, the

argument for this claim, to be developed in Chapter 6, differs substantially from that of Regan. In fact, I find myself unable to accept Regan's theory, not, or not primarily, because of the content of the rights it adduces and defends, but because of their metaphysical basis. The first part of this chapter presents an overview of Regan's theory, with particular reference to its metaphysical basis: the concepts of a subject-of-a-life and inherent value. The final sections offer a critique of Regan's position, again with particular reference to its reliance on these concepts. This sets up discussion of Chapter 6, where an alternative account of moral rights is developed.

1 Subjects-of-a-life

The conceptual edifice upon which Regan's rights-based view is built is composed of two concepts: *subject-of-a-life* and *inherent value*. This section deals with the notion of a subject-of-a-life, the following with inherent value. According to Regan, an individual is a subject-of-a-life if it possesses the following sorts of features:

Individuals are subjects-of-a-life if they have beliefs and desires; perception, memory, and a sense of the future, including their own future; an emotional life together with feelings of pleasure and pain; preference and welfare-interests; the ability to initiate action in pursuit of their desires and goals; a psychophysical identity over time; and an individual welfare in the sense that their experiential life fares well or ill for them, logically independently of their utility for others and logically independently of their being the object of anyone else's interests.¹

These conditions are collectively referred to as the *subject-of-a-life criterion*. Many creatures, according to Regan, satisfy these conditions. The conditions are, for Regan, certainly satisfied by almost all humans, including young children and the mentally enfeebled. Living human beings in persistent vegetative states might not satisfy this criterion, however, and humans in irreversible coma

almost certainly would not. The criterion is also satisfied by all normal members of mammalian species (the exceptions would, presumably, be analogous to the human exceptions listed above). It is also likely to be satisfied by many species of birds, and quite possibly by reptiles, amphibians, and fish, although Regan does not wish to take a stand on these latter types of case. And, thus, his eventual case for animal rights will, strictly speaking, be restricted to mammals. Regan will, in effect, present a case for mammalian rights. The question of whether birds, reptiles, amphibians, and fish satisfy the subject-of-a-life criterion is, ultimately, an empirical one; but, if it should turn out that they do, then Regan's case can easily be extended to include them. It is primarily to avoid any controversial, or at least questionable, empirical assumptions that Regan restricts his arguments to mammals.

What is crucial to the role that the subject-of-a-life criterion will play in Regan's argument is that it cuts across the distinction between moral agents and moral patients. Both can be subjects-of-a-life. This, Regan argues, is a (morally) relevant similarity between moral agents and patients, a similarity that he will exploit in the arguments to follow.

It should be noted that there are two different ways in which the subject-of-a-life criterion might be understood. According to what I shall call the *strong* interpretation, an individual must satisfy all of the conditions listed above in order to qualify as a subject-of-a-life. That is, the conditions collectively constitute a set of necessary and sufficient conditions for something being a subject-of-a-life. In order to be a subject-of-a-life, you must possess *all* the features listed in Regan's description. According to the *weak* interpretation, the constraints imposed by the features on Regan's list are somewhat softer. On this view, in order to be a subject-of-a-life you must satisfy *most* of the above conditions, but not necessarily all. And which conditions must be satisfied can vary from case to case. Thus, on the weak interpretation, one creature might qualify as a subject-of-a-life because it satisfies

the conditions of perception, belief, desire, memory, emotional life, psychophysical identity over time, and so on, but does not have a sense of the future. Another might have a sense of the future (e.g., it can anticipate), but no clear psychophysical identity over time. On this weaker view, then, creatures satisfying the subject-of-a-life criterion need bear only what Wittgenstein has called a relation of *family resemblance* to each other. They need not share precisely the same features. Or, to employ somewhat more up-to-date jargon (borrowed from artificial intelligence), the conditions listed under the subject-of-a-life criterion function as *soft constraints* – conditions which are significant, but not necessarily of overriding importance.

Regan is not explicit on whether he advances the subject-of-a-life criterion in the strong or weak sense. However, the context provided by his later arguments suggests that he intends the strong interpretation of the criterion. This, as we shall see later in the chapter, might well be a mistake. It leaves Regan's argument open to essentially irrelevant objections.

2 Inherent value

According to Regan, all creatures which are subjects-of-a-life have *inherent value*. Being a subject-of-a-life is a *sufficient* condition for having inherent value, not a *necessary* one. That is, if you are a subject-of-a-life, then you have inherent value, but if you are not a subject-of-a-life, it does not necessarily mean that you don't have inherent value. It is, therefore, possible that humans and animals who don't meet the subject-of-a-life criterion nonetheless do have inherent value.

There are four features central to the concept of inherent value. Firstly, the inherent value possessed by an individual is independent of their being the object of anyone else's interests. Possession of inherent value does not depend on whether, or how much, one

is liked, respected, or in any other way valued by others. As Regan puts it, the lonely, forsaken, unwanted, and unloved have no less (and no more) inherent value than those in more fortuitous social circumstances.

Secondly, the inherent value of an individual does not vary according to the extent to which they have utility *vis-à-vis* the interests of others. The unscrupulous used-car salesman has just as much inherent value as the most beneficent philanthropist.

Thirdly, the inherent value of an individual is not something they can earn or cultivate by dint of their efforts; and it is not something they can lose by what they do or fail to do. A criminal, no matter what his crime, is no less inherently valuable than a saint.

Fourthly, and, for Regan's purposes, perhaps most importantly, the inherent value of an individual is conceptually distinct from, and not reducible to, whatever value attaches to the experiences had by that individual (Regan refers to this as *intrinsic* value). It is not possible to determine the inherent value of an individual by totalling up the intrinsic values of their experiences. Inherent value is simply *incommensurable* with intrinsic value: the two simply cannot be compared; they cannot be assessed by the same scale of measurement. Thus, those who have a generally happy life do not have more inherent value than those who do not, even though they presumably undergo more experiences with intrinsic value. Nor do those with more 'sophisticated' or 'cultivated' preferences have more inherent value than those whose pleasures tend towards the vulgar or earthy. The inherent value of an individual is something which is distinct from, not reducible to, and actually incommensurate with the values of those experiences which that individual undergoes during the course of its life.

The notion of inherent value thus allows Regan to draw a clear distinction between his view and utilitarianism. Utilitarianism, for Regan, treats individuals as *receptacles* of value.² For utilitarians, the primary locus of value lies in the experiences which a person undergoes (usually pleasures or preference-satisfactions). An

individual human or animal, on this view, is like a cup which, in itself, has no value but, circumstances permitting, can become the container of valuable things – pleasures or preference-satisfactions, depending on the type of utilitarianism in question. For Regan, on the other hand, an individual is like a cup which has value in itself, that is, inherent value. This cup can contain things that are valuable – for example, certain sorts of experiences – but the value of the cup is distinct from, not reducible to, and cannot even be compared with, the value of these things that it contains. Individuals who satisfy the subject-of-a-life criterion have, according to Regan, this kind of inherent value.

Thus, according to Regan, the *inherent* value of an individual subject-of-a-life is incommensurate with the *intrinsic* value of that subject's experiences or other mental states. Therefore, Regan argues, the former can never be overridden or outweighed by the latter. It is not legitimate, then, in a sense to be made clear, to justify a situation which involves riding roughshod over someone's inherent value merely by appealing to the more favourable aggregation of intrinsic value that this situation involves or produces. This is not to say, however, that the intrinsic value of experiences is irrelevant to moral decision-making. As we shall see, there are, even on Regan's account, situations in which such value clearly is relevant. It does mean, however, that inherent value is a sort of moral *trump* with respect to the intrinsic value of experiences. One cannot legitimately override the inherent value of an individual by appeal to the value of the experiences, either in that individual or in others, which this would bring about.

3 Inherent value as a theoretical postulate

It seems, at least at first glance, that Regan's argument is a straightforward version of the so called *naturalistic fallacy*; very roughly, the fallacy (if indeed it is a fallacy) of inferring values from facts.

That is, it seems as if Regan has simply presented us with an argument of the following form:

Premise. Creature X is a subject-of-a-life
Conclusion. Creature X has inherent value

To view Regan's argument in this way, however, would be to seriously misrepresent it. The argument for inherent value is not a version of the naturalistic fallacy, but, rather, has the form of an *inference to the best explanation*.

An inference to the best explanation has the following form. We start off with a phenomenon, or set of phenomena, which need explaining. We then hypothesize the existence of a certain entity (or, in some cases, law or principle) which is capable of explaining that phenomenon. Finally, it is argued that the hypothesized entity is the most plausible explanation of the phenomenon because all competing explanations are manifestly, or at least arguably, false. This is essentially the type of argument Regan is giving for inherent value.

The phenomenon which needs explaining in this case is what Regan calls our *considered beliefs* about moral issues. In particular, one of our considered beliefs about morality, indeed perhaps the most fundamental, is that we have a duty to treat people justly. Treating people with justice is not an optional or *supererogatory* moral principle, it is essential to the nature of morality as such. Moreover, our considered moral beliefs also incorporate a fairly definite conception of what is involved in treating someone with justice. Thus, Regan points out that our considered moral beliefs rule out a perfectionist view of justice, according to which what individuals are due, as a matter of justice, depends on the degree to which they possess a certain cluster of virtues or excellences. These might include intellectual and artistic talents, and/or a certain sort of character. According to perfectionist views of justice, individuals who possess these virtues in abundance are due more, as a matter of justice, than those who do not. This sort of view is,

as Regan points out, morally pernicious, providing a justification for seriously objectionable forms of social and political discrimination – slavery, caste systems, and so on. And these forms of discrimination, our considered moral beliefs inform us, should be rejected. Thus, perfectionist views of justice do not cohere with our considered moral beliefs.

Utilitarian accounts of justice also do not seem to cohere very well with our considered beliefs about justice. As we have seen, standard objections to utilitarianism point out that it seems, in principle, to legitimize many forms of what we would regard as serious injustice, as long as the overall aggregate of pleasure or preference-satisfaction is increased. Thus, we would be justified in treating an individual with what we would intuitively regard as injustice, as long as doing so secured a greater amount of happiness or preference-satisfaction in the world. In other words, innocent individuals could, in principle, be sacrificed for the greater good of the community. And, at least intuitively, our considered moral beliefs tell us that such a situation is a paradigm case of injustice.

Regan believes that in order to account for our considered beliefs about justice, we must postulate that certain sorts of individuals – individuals who are subjects-of-a-life – possess inherent value in the sense explained above. If we suppose that subjects-of-a-life do possess inherent value, then we can explain our considered beliefs about the importance of justice and about what just treatment amounts to. So, the postulation of inherent value, at least on Regan's view, is an explanation of our considered moral beliefs – or as Rawls would put it, our *reflective intuitions* – concerning justice. And given the manifest, or at least arguable, failure of other theories to account for these beliefs or intuitions, we have good reason for supposing that the postulation of inherent value is the *best* explanation of these beliefs. Thus, inherent value is a theoretical postulate, justified, on Regan's view, as an inference to the best explanation. It is no different in kind, he would claim, to the postulation of, for example, atomic particles to explain characteristic patterns in a

cloud chamber; or the postulation of an additional planet to explain perturbations in the orbit of Neptune.

4 The respect principle

Once we allow that all subjects-of-a-life have inherent value, we can derive several important moral principles concerning how they should be treated. The first of these Regan calls the *respect principle*:

We are to treat those individuals who have inherent value in ways that respect their inherent value.

That is, all individuals who are subjects-of-lives must, as a matter of justice, be treated in ways that respect this fact. The respect principle sets forth an egalitarian, anti-perfectionist interpretation of the formal principle of justice. Treating an individual that possesses inherent value in ways, and only in ways, that respect this value is not supererogatory, not an optional moral extra. It is required by justice; any contrary treatment is unjust.

We fail to treat individuals who have inherent value with the respect they are due whenever we treat them as if they lacked inherent value. And we treat them in this way whenever we treat them, as the utilitarian does, as if they were mere receptacles of valuable experiences such as pleasures or preference-satisfactions. We also treat them as if they lacked inherent value when we treat them as if their value depended upon their utility relative to the interests of others. And we also treat them as if they lacked inherent value when we harm them simply so that we may bring about the best aggregate consequences for everyone affected by the outcome of such treatment. All these are, in fact, just variations on the same theme: treating an individual with inherent value as if he, she or it were nothing more than a receptacle of value. That is, it is to treat something with inherent value as if it lacked inherent value; and this is unjust.

The respect principle, however, requires more than that we abjure from treating an inherently valuable individual in disrespectful ways. The principle, in fact, imposes on us a *prima facie* duty to assist those who are the victims of disrespectful treatment (i.e., injustice) at the hands of others. This is, in fact, a common view of justice, and not peculiar to Regan's account. That is, it is commonly accepted that justice, whatever its interpretation, not only imposes duties of non-harm; it also places us under a *prima facie* obligation to aid those who are the victims of injustice. As Regan puts it, all those who have inherent value are to be given what, as a matter of justice, they are due; and sometimes what they are due is our assistance.

The respect principle, for Regan, may or may not be morally fundamental in that it may or may not be derivable from a more fundamental principle. Regan leaves the question of its ultimate logical status open. The justification for the principle, however, is essentially the same as the justification for the postulation of inherent value. That is, no moral theory which fails to incorporate the respect principle can hope to systematize, justify, and, above all cohere with, our considered beliefs about justice.³

5 The harm principle

Unlike the respect principle, the *harm principle* is not a candidate for basic moral principle: it can, in fact, be derived from the respect principle. The harm principle states that:

We have a direct *prima facie* duty not to harm individuals.

According to the respect principle, any individual who has inherent value is owed, as a matter of strict justice, treatment that is respectful of this value. And this amounts to the claim that any individual who is a subject-of-a-life is owed, again as a matter of justice, treatment which respects this fact. Now, any individual

who is a subject-of-a-life has an *experiential welfare*; that is, their life can, from their perspective, fare well or ill for them. Their life, that is, can go experientially better or worse for them, logically independently of their utility for others and of their being the object of another's interests. Therefore, the concepts of benefit and harm apply to these sorts of beings in virtue of the fact that they are subjects-of-a-life. That is, being the subject-of-a-life bestows on an individual the distinctive sort of value which consists in having an experiential welfare. Therefore, at least *prima facie*, we fail to treat individuals in ways that respect their value when we treat them in ways that detract from their welfare. And we detract from the welfare of this type of an individual when we harm them. Therefore, Regan claims, we have a *prima facie* direct duty not to harm those individuals who have an experiential welfare. And this is precisely what the harm principle claims. Thus, the harm principle is derivable from the respect principle.

According to Regan, therefore, we have a *prima facie* duty not to harm those individuals who are subjects-of-a-life. The qualification *prima facie* signifies that the duty can, in certain circumstances, be overridden. That is, the harm principle does *not* entail that, no matter what the circumstances, it is always wrong to harm an individual who is the subject-of-a-life. As we shall see, there are, according to Regan, circumstances when harming an individual with inherent value, even an innocent individual, is morally legitimate. This issue will be explored in later sections.

6 Moral rights

Justice requires, then, that we should treat those individuals who possess inherent value – that is, who are subjects-of-a-life – in accordance with the respect principle and the harm principle. This provides the basis for Regan's claim that individuals with inherent value also possess moral rights. Regan's demonstration of this latter claim proceeds by way of an analysis of the concept of a right.

Regan adopts the widely accepted view of moral rights as *valid claims*.⁴ The relevant sort of valid claims have two aspects, (i) a valid claim-to, and (ii) a valid claim-against.

A claim-to, in this context, is a claim to a certain commodity, freedom, or type of treatment by others. And to be a valid claim-to, the claim must be backed, or validated, by an appeal to a correct moral principle or principles. Thus, the claimant can demonstrate that she is owed the commodity, freedom, or treatment in question by appeal to the relevant moral principles.

In order to be a valid claim-against, a claim must be made against assignable individuals who do in fact owe what the claimant asserts. And, again, whether the individuals in question do owe the commodity or treatment must again be decided by appeal to correct moral principles.

When both a claim-to and a claim-against has been validated by appeal to correct moral principles, we can speak of a *valid claim all things considered*. And this, according to the present analysis, is what constitutes a moral right. To have a moral right to a certain commodity or treatment is to have a valid claim all things considered to that commodity or treatment, and a valid claim all things considered against whatever individuals are to provide the commodity or treatment.

Regan has, of course, already argued that the respect principle and the harm principle are valid moral principles. Therefore, a claim, made against assignable individuals, to a certain commodity or treatment will be a valid claim all things considered if it is backed or validated by an appeal to either the respect or the harm principle. Therefore, any individual with inherent value has a moral right to treatment that respects this value. The right to such treatment is a valid claim-against assignable individuals (i.e., all moral agents) and a valid claim-to a certain type of treatment, the validity of each claim being backed by the respect principle, a valid moral principle. Similarly, any individual with inherent value has a *prima facie* right not to be harmed. Such a right is again a valid claim-to and a valid claim-against, the validity of each being backed by the

harm principle, a correct moral principle. Therefore, any individual who has inherent value (i.e., a subject-of-a-life) has a moral right to treatment that respects this value and a *prima facie* moral right not to be harmed.

Two points should be noted. Firstly, the right not to be harmed is a *prima facie* moral right only. That is, it can be overridden in certain circumstances (to be clarified shortly). The *prima facie* status of the right, here, is due to the *prima facie* status of the harm principle from which it derives.

Secondly, according to Regan, one consequence of this analysis is that one can have moral rights only against moral agents; not against moral patients or inanimate objects. The reason for this stems from the nature of claims-against. A claim-against can be a valid one only if the individual against whom the claim is made is capable of meeting the requirements of the claim. That is, the individual must be capable of providing the treatment or commodity that, according to the claim, is due. Thus, for example, we can have no moral rights against nature. We would have such rights only if nature was capable of providing us with the commodity or treatment we claimed was due. But nature is obviously incapable of acting in such a way. More precisely, we could have valid claims-against nature only if nature has direct duties to us to do or forbear doing certain acts that are our due. But nature, in this sense, is capable neither of doing nor forbearing from doing things. To say that nature *ought* to do certain things, or forbear from doing certain things, is to presuppose that nature *can* choose in the relevant sort of way. But nature obviously cannot choose what it does. Therefore, we have no rights against nature because we have no valid claims-against nature.⁵

This also allows us to dispense very neatly with a frequently raised objection to the concept of animal rights: the claim that the concept leads to absurdity.⁶ The argument runs as follows. If sheep have rights, then these are violated by wolves that prey on them. Therefore, it seems that if we have a duty of assistance to sheep to stop those who violate their rights, then it seems we have a duty to

stop wolves from preying on sheep. However, if we were to do this, we would be violating the wolves' rights by harming them (e.g., by consigning them to a slow painful death through starvation). Thus, either way we end up violating some creature's rights. And, therefore, the whole concept of animal rights leads to logical absurdity.

Once, however, we understand that moral rights are valid claims, this objection can be stopped, so to speak, before it even starts. The sheep has no moral rights with respect to wolves. That is, the sheep does not have a valid claim-against the wolf to refrain from eating it. This is because the wolf is not a moral agent, hence is not capable of choosing, in any morally relevant sense, whether or not to eat the sheep. The sheep would have a valid claim-against the wolf in this regard only if the wolf had the capacity to forbear from eating the sheep. And the wolf has no such capacity. That is, to say that the sheep has a right against the wolf not to eat it is to imply that the wolf ought not to eat the sheep. And to say that the wolf ought not to eat the sheep is to imply that the wolf is capable of choosing whether or not to eat the sheep. But the wolf has no such capacity. Therefore, the sheep has no valid claim-against the wolf not to eat it. And, therefore, the sheep has no moral right against the wolf in this regard. Neither, for that matter, do humans (not that wolves ever eat humans). It makes no sense to speak simply of a moral right to X as such. To speak of a moral right is always an elliptical way of referring to the individual against whom the right is claimed; it presupposes such an individual or individuals. And we can only have rights against moral agents.

7 The miniride principle

According to Regan, the moral right not to be harmed, possessed by an individual with inherent value, is a *prima facie* moral right. That is, the right can, in certain circumstances, be overridden. The task Regan now has, then, is to show in a *principled* way what these circumstances are. That is, he does not just want to claim, in an

ad hoc fashion, that the right can be overturned in circumstances X. In order to make the exceptions theoretically satisfactory, he must show how they can be derived from the respect or harm principles themselves. If this cannot be done, then the exceptions, in effect, count as objections to Regan's theory. However, if the exceptions can be derived from Regan's theory itself, then they count not as objections to the theory but, in an importance sense, as confirmations of it. Therefore, Regan describes two principles which govern the circumstances under which the harm principle can be overridden, and tries to show how these can be derived from the respect principle.

The first of these is what Regan calls the *miniride principle* (the minimize overriding principle):

Special considerations aside, when we must choose between overriding the rights of many who are innocent or the rights of few who are innocent, and when each affected individual will be harmed in a *prima facie* comparable way, then we ought to choose to override the rights of the few in preference to overriding the rights of the many.⁷

This principle, Regan argues, is derivable from the respect principle. The derivation runs as follows.

The respect principle entails that all individuals with inherent value have a *prima facie* right not to be harmed, and all those who have this right have it equally. Therefore, precisely because this right is equal, no one individual's right can count for more than the right of another, at least when the harm that will befall both is *prima facie* comparable. But, therefore, for any individual with inherent value, precisely because each individual's possession of the right not to be harmed is equal to the right of every other individual, one should, in a situation where one is forced to choose between overriding the right not to be harmed of few and overriding the right not to be harmed of many, choose to override the right not to be harmed of the few. To choose otherwise would be to accord inordinate status

to the rights of the few. That is, it would be to imbue the rights of the few with greater value or significance than the same rights of the many. And this is precisely what the respect principle says you should not do. To choose in favour of the few, that is, would involve not treating the many with the respect that they deserve as bearers of inherent value. But to choose in favour of the many does not entail that one fails to treat the few with the respect they deserve. In this way, therefore, the miniride principle is derivable from the respect principle.

The miniride principle, as the above formulation makes clear, has two qualifications. The principle is restricted to cases where the harm suffered by each inherently valuable individual is *prima facie comparable*. And the principle is restricted to situations where no *special considerations* obtain. Consider an example to illustrate each of these in turn.⁸

Fifty-one miners are trapped in a mine cave-in and are certain to die in a very short time if nothing is done. Fifty of the trapped men are located on the pit floor, and the other one is located in a shaft leading down to the floor. Suppose the only one way to reach the fifty miners in time is to place an explosive charge in the shaft through which the trapped men can then escape. However, suppose also that this method is certain to kill the lone miner. The one miner could be saved, however, if we simply dug him out, but the time this would take would certainly lead to the death of the 50 trapped men. What ought we to do? In this case, the miniride principle says that we should save the 50 men at the expense of the one.

In the above case, the harm facing each of the trapped men is comparable – all 51 of them stand to die if nothing is done. However, suppose the potential harm facing the two groups was not comparable. Suppose there was a way of getting the single miner out while also saving the lives of the 50. Suppose, for example, that we did have time to dig the single man out, but that this would lead to substantial delay in rescuing the other miners. We have, however, every reason to think that the mine has stabilized and that there would be no further cave-ins, nor is there any danger of explosions,

and so on. We know that several of the group of 50 miners have suffered painful injuries – broken legs, and the like; but that none of these injuries are life-threatening. We thus have a choice between sacrificing the life of the lone man, or allowing the group of 50 to remain for, say, 48 hours in a state of quite significant pain coupled with the fear of the mine collapsing at any point. Now, in this scenario, what ought we to do? Well, whatever we ought to do in this case, the miniride principle does not apply. The reason is that the harm suffered by the single miner and that suffered by the group of 50 is not, *prima facie*, comparable. The harm facing the lone miner is death; that facing the group of 50 is pain and fear. The harms are not comparable, and, therefore, the miniride principle does not apply in this case.

Consider, now, the notion of special considerations. Suppose the single trapped man had been kidnapped by the group of 50, who hoped to reap the financial rewards of his forced labour. This would be a special consideration, and, once again, the miniride principle would not apply in this case. Or perhaps the group of 50 had, for some difficult to imagine reason, each signed a legal document requesting that in the event of this sort of situation, person X should be saved before them. And suppose the lone miner was person X (this is a difficult situation to imagine, I know, but the point concerns not the plausibility of the scenario but the principle underlying it). There are, in fact, several different types of special consideration Regan is prepared to allow as morally significant, but the details do not concern us here. The point is that, as soon as special considerations obtain, the miniride principle no longer applies.

8 The worse-off principle

The second principle determining when the harm principle may justifiably be overridden is what Regan calls the *worse-off principle*:

Special considerations aside, when we must decide to override the rights of the many or the rights of the few who are innocent,

and when the harm faced by the few would make them worse-off than any of the many would be if the other option were chosen, then we ought to override the rights of the many.⁹

Although the principle is formulated in terms of the rights of the many and the few, such formulation is not essential to the principle as such. If we were forced to choose between harming *one* innocent individual and harming another, the worse-off principle would still apply and legislate in favour of the individual who would be made worse-off.

The worse-off principle is also derivable from the respect principle. The derivation goes like this. Suppose we have two individuals, P_1 and P_2 , both of whom have inherent value. The respect principle entails that P_1 and P_2 have an equal right not to be harmed, a right which derives from the equal inherent value possessed by each. However, despite the fact that they possess an equal right not to be harmed, this does not entail that each and every harm either may suffer is equally harmful. All things being equal, P_1 's death is a greater harm than P_2 's (non-fatal dose of) flu, even if both possess an equal right not to be harmed. But this means that in order to show equal respect for the equal rights of the two, one must count their equal harms equally; one must not count their unequal harms equally. If we were to count unequal harms equally, this would imply that we were not, in fact, according due respect to the equal rights of each individual. To attempt to alleviate P_2 's flu, at the expense of P_1 's death would be to give P_2 more than his due. P_1 and P_2 , as inherently valuable individuals, have an equal right to respect, and, consequently, an equal *prima facie* right not to be harmed. And precisely because of this, and because the harm P_1 faces is greater than the harm faced by P_2 , equal respect for the two requires that we not choose to override the right of P_1 but choose, instead, to override that of P_2 .

Now, according to Regan, adding numbers in this case makes no difference. Suppose we have to weigh the death of P_1 against the (non-fatal) flu suffered by $P_2, P_3, \dots, P_{1000}$. The death of P_1 would, according to Regan, still be a greater harm because it is

greater than the individual harm suffered by each and every of the remaining 999. There is, after all, no aggregate individual who suffers the sum of the harms suffered by the remaining 999 individuals. There are just the 999 individuals, none of whom will be worse-off than P_1 would be. It is the magnitude of the harm done to P_1 and each individual member of the remaining 999, not the sum of P_1 's harm compared with the sum of the harms done to the 999 that determines whose right overrides whose. Thus, according to Regan, since the harm done to P_1 would be greater than that done to any other individual, and would make P_1 worse-off than any other individual involved, respect for the equal rights of everyone involved requires overriding the rights of the many rather than those of the individual. In general, in the absence of special considerations, if the few who are innocent would be made worse-off than any of the many who are innocent, the respect principle requires that we override the rights of the many. And this is precisely what the worse-off principle states. Therefore, the worse-off principle is derivable from the respect principle.

It might be thought that, in putting forward this argument, Regan is undermining his own rights-based position. After all, the rights view is supposed to deny the moral relevance of consequences; this is the basis of its opposition to utilitarianism. But now Regan seems to be relying on the notion of comparable harm and invokes considerations about who will be harmed most. Thus, in defending the worse-off principle, Regan's position seems to be inconsistent.

However, there is, in fact, nothing in Regan's position which commits him to the claim that consequences are irrelevant to moral decision-making. What the rights view *does* deny is that moral decisions can legitimately be made *merely* by determining which alternative will bring about the best aggregate consequences for all those affected by the decision. Consequences, in other words, are relevant; it's just that they are not the only relevant factor. The rights view, therefore, does *not* claim that consequences are morally irrelevant. *A fortiori*, it does *not* claim that we can dispense with consideration of consequences in making our moral decisions and determinations. Indeed, it would be a very strange, and

extremely implausible, view if it did make these claims. The rights view, in fact, entails that the consequences of actions *are* extremely relevant considerations in moral decision-making. It would not be possible to show equal respect toward each of the individuals affected by an action if we did not weigh up how much each would be harmed by that action. Therefore, the rights view does not entail that consequences in general, and consequences for the specific welfare of individuals in particular, are morally irrelevant. In fact, when the rights view is properly understood, it entails just the opposite. It denies only that consequences are the *sole* basis upon which moral decisions should be reached.

The miniride and worse-off principles can be, roughly, summed up in the following slogan: special considerations aside, when the harms are comparable numbers count, when the harms are not comparable numbers don't count. Both of these principles should be clearly distinguished from a third one which Regan does not endorse. This is what Regan refers to as the minimize harm principle: act so as to minimize the total aggregate of harm of the innocent. The minimize harm principle is a purely consequentialist principle. It instructs us to act so as to avoid the worst consequences, where these are understood as the greatest sum of harm done to all the innocents affected by the outcome. And to accept this principle, therefore, is to assume that inherently valuable individuals are mere receptacles of value. They are receptacles, in this case, not of pleasures and pains but of harms and benefits. And the minimize harm principle tells us that we should minimize the total amount of harm irrespective of what inherently valuable individuals we have to sacrifice toward this end. Both the miniride and worse-off principles, therefore, should be clearly distinguished from the minimize harm principle.

9 Regan on vegetarianism

So far, then, Regan has used the notion of a subject-of-a-life to motivate and defend the concept of inherent value. He has used

the concept of inherent value to justify the respect principle, and from the respect principle he derived the harm principle. These, being valid moral principles, were used to defend the idea that all inherently valuable individuals have a moral right to be treated in ways which respect this value, and a *prima facie* moral right not to be harmed. Exceptions to the right not to be harmed are governed by the miniride and worse-off principles, both of which, Regan argues, can be derived from the respect principle. In short, what Regan has provided here is a consistent, cogent, and systematic moral theory which not only attempts to justify the concept of a moral right but also to determine priority rules for the moral rights thus justified. It is an elegant theory, and this is never more evident than in its application to particular moral issues. This section tries to provide a feel for the ways in which Regan's theory has application to the world. The particular moral issue we shall look at is the raising, killing, and eating of animals (specifically, mammals) for food.

Vegetarianism, according to Regan, is morally obligatory. That is, it is not simply a morally good thing to be a vegetarian; it is a morally bad thing not to be a vegetarian. The basic reason for this should be quite evident from his overall theory. In fact, Regan's basic argument for vegetarianism looks something like this.

- P1. Mammals are subjects-of-a-life.
- P2. As such, mammals have inherent value.
- P3. Therefore, the respect principle, and the harm principle derivable from it, apply to mammals.
- P4. Therefore, mammals have a right to be treated with respect, and a *prima facie* right not to be harmed.
- P5. Raising, killing, and eating of mammals harms them.
- C. Therefore, we should not raise, kill, and eat mammals.

This, in any event, is the most obvious way in which Regan's theory applies to the issue of raising, killing, and eating of mammals. The restriction to mammals is because of the earlier

mentioned restriction on the scope of the subject-of-a-life criterion. It is fairly clear, Regan thinks, that the criterion applies to mammals; but its application to non-mammals is more controversial. And in wishing to keep the premises of his argument as non-controversial, and therefore as widely acceptable, as possible, Regan restricts his case to mammals.

However, this is not the end of the argument. The right not to be harmed is, after all, a *prima facie* right; a right that can be legitimately overridden in certain circumstances. Any support for the raising, killing, and eating of mammals, therefore, will likely focus on the *prima facie* status of the right not to be harmed, and try to show that, in this case, the right can be overridden for legitimate moral reasons. And this, in effect, is where the hard work begins.

The claim that the *prima facie* right of non-human mammals not to be harmed can be overridden to allow us to use them for food is often thought to gain support from another principle; a principle which is also derivable from the respect principle. Regan refers to this as the *liberty principle*:

Provided that all those involved are treated with respect, and assuming that no special considerations obtain, any innocent individual has the right to act to avoid being made worse-off even if doing so harms other innocents.¹⁰

The liberty principle is derivable from the respect principle. As an individual with inherent value, I am always to be treated with respect and thus am never to be viewed or treated as a mere receptacle or as one who has value merely relative to the interests of others. Furthermore, I also have a welfare, and I should be allowed to do whatever is necessary to advance this welfare, as long as I treat others with respect. To deny me the liberty to pursue my welfare simply because others will be worse-off if I do is to fail to treat me with the respect that I, as a possessor of inherent value, am due, as a matter of justice. Thus, for example, I should be able to compete for a job with another even if my success means that

he is destined for the poor house. He will be made worse-off, and undoubtedly harmed, by my success, but to deny me the liberty to pursue the job simply because of this is to give me less than my due: it is to assume that the treatment I am due, as a matter of justice, is contingent upon how others will be affected as a result. It is to fail to give my inherent value proper consideration. The qualification of 'special considerations' is meant to handle such cases as where, for example, I steal my neighbour's Mercedes on the grounds that not to do so would be to make me worse-off relative to him. The liberty principle would not permit this action, on the grounds that the ownership of the car by my neighbour is a special consideration and hence is not covered by the liberty principle.

The second proviso – namely 'provided that all those involved are treated with respect' – is more complicated, and has an extremely important consequence. Suppose I were a sadist whose greatest pleasure was the torturing of innocents. Then, if I were not able to indulge my passion in this regard, I would, at least arguably, be made worse-off. However, to engage in my interest in torture would, manifestly, not respect the inherent value of my victims. What this shows is that it is only possible to respect the inherent value of those affected by my actions if the standard of me being made worse-off is set relative to them. That is, according to the liberty principle, I can legitimately harm innocents only if this is to avoid being made worse-off *than them*. That is, *it is not being made worse-off as such which is crucial, but being made worse-off than those innocents I must harm*. That is, the liberty principle only justifies overriding the harm principle when I must act to avoid being made worse-off than those I must harm in the process. Since it is hard to see how by not torturing innocents I would thereby be made worse-off than they would be if I had tortured them, the liberty principle rules out this sort of action. This is an extremely important qualification. Without it, it is difficult to see how any defence of vegetarianism could get off the ground; since it is at least arguable that we are made worse-off by not eating meat. The crucial point is not whether we are made worse-off, but whether, by not eating

meat, we are thereby made worse-off than the animals we raise and kill would be if we were to eat meat.

Therefore, in addition to the respect, the harm, the miniride, and the worse-off principles, then, the rights view recognizes a fifth principle, the liberty principle, to be understood in accordance with the above qualifications. And opponents of Regan's claim that vegetarianism is morally obligatory are likely to appeal to this principle. For since both farmers and meat-eaters are authorized to act as the liberty principle allows, they may claim that they are at liberty to raise and eat animals, even though this involves harming them, because not to do so would make them worse-off relative to any of those individuals who are harmed in the process – that is, relative to any farm animal. Moreover, Regan's defence of the worse-off principle has already ruled out aggregating the harms done to farm animals in order to defend vegetarianism.¹¹ But what, however, could be the grounds for claiming that the harm done to farm animals is justified by the liberty principle? The following by no means constitutes an exhaustive treatment of Regan's argument for vegetarianism. I have chosen, instead, to focus on those arguments which might be thought to constitute the biggest threat to Regan's position. That is, I have chosen to focus on those arguments, used to defend the practice of animal husbandry, which might, *prima facie*, be thought to be justified by appeal to the very principles Regan himself defends. Focusing on these arguments will allow us to see more clearly the subtlety and power of Regan's position.

One argument commonly raised in favour of animal husbandry focuses on the pleasure human beings get from both eating and preparing meat dishes. Animal flesh, it is argued, tastes very nice, and to abstain from eating it is to forgo certain pleasures of the palate. Moreover, from a culinary standpoint, it is personally rewarding to prepare dishes of this nature. Therefore, should human beings choose to forgo eating meat, they would be clearly worse-off from being denied these sorts of pleasures. Therefore, it might be argued, the liberty principle itself licenses the continuing practice of animal husbandry: it is legitimate for us to continue with this practice

since not to do so would make us worse-off than we would otherwise have been.

There are several fairly obvious problems with this argument. Firstly, no one has a right to eat something just because they happen to find it tasty. If I happen to find babies good to eat, it does not mean that I have the right to eat them. Nor does it mean that if I find the cooking of babies to be a rewarding culinary adventure, I thereby have the right to cook them. Secondly, there are many other tasty foods besides meat, and other foods beside meat can offer similar, or greater, culinary rewards. So, it is not even clear that we are harmed by not being able to eat meat. But thirdly, and most importantly, even supposing we were harmed (i.e., made worse-off) by not eating meat, the harms we would be called upon to endure could not reasonably be viewed as *prima facie* comparable to the harm visited upon farm animals. That is, even if we were made worse-off by not eating meat, we would be nowhere near as badly off as farm animals would be were we to eat meat. Our being made worse-off by forgoing certain pleasures of the palate or kitchen can, by no stretch of the imagination, be compared to the daily suffering, deprivation, and untimely death of farm animals. Now, the liberty principle, remember, does not claim that we are morally entitled to harm innocents in order to avoid being made worse-off. It says that we are entitled to do so *if* all those involved are treated with respect. And this amounts to the claim that we are entitled to harm innocents to avoid being made worse-off only if not doing so would make us worse-off than the innocents would be if we chose to harm them. Therefore, the liberty principle does not, at least in this case, justify our consumption of meat, since, by forgoing cooking and eating meat we would not be made anywhere near as badly off as farm animals are made by our not forgoing these activities. Moreover, as Regan points out, matters do not substantially change if the farm animals in question are raised and slaughtered 'humanely'. Death is one of the greatest harms that can be inflicted on an inherently valuable individual, and one cannot compare the harm of death with the harm of abstaining from certain adventures

of the kitchen or palate. In fact, once we get clear on the principles involved here, we see that animal husbandry is ruled out by the worse-off principle. The death and suffering inflicted on animals by this practice makes them substantially worse-off than we would be made if we chose not to, or were prevented from, eating meat.

Another common argument in favour of animal husbandry focuses on economic considerations. Some people, it is argued, have a strong economic interest in continuing to raise farm animals, and the quality of their life as well as that of their dependants is tied to the continuation of this practice. This claim might be thought to be backed by both the worse-off principle and the liberty principle. Consider, first, the worse-off principle. The farmer, it might be claimed, will be made worse-off relative to the animals he raises if we, the consumers, became vegetarians and thereby failed to support him. He would lose his livelihood, and, with this loss, the harm done to him would outweigh any harm done to any one animal, even one kept in close confinement. And remember, on the rights view, we cannot aggregate the lesser *prima facie* harms of the many as a way of justifying causing greater *prima facie* harm to the few. Therefore, we ought to eat meat because we owe it to the farmer – this is what he is due as a matter of justice. And the principle backing this claim is the worse-off principle.

However, proper understanding of the worse-off principle, Regan argues, reveals that this is, in fact, not the case. The basic reason for this is that to enter into any business is to run the risk of failure. It is also to acknowledge, at least tacitly, both that no one has a duty to purchase one's products or services and that such purchase cannot be claimed as one's due. Voluntary participation in a competitive activity such as a business constitutes a *special consideration*. And the worse-off principle is explicitly formulated to begin 'Special considerations aside'. Therefore, the worse-off principle cannot legitimately be applied in this case. As Regan points out, a businessperson who would be made worse-off if her products or services were not purchased has no valid claim on anyone to make the necessary purchases to keep her business afloat.

Business is a competitive activity, and those who voluntarily participate must understand that the worse-off principle is suspended. We do not, therefore, *owe* it to the farmer who sells meat to purchase his products.

Another argument relates the farmer's economic interests to the liberty principle. Even if we have no duty to purchase the farmer's products, it could be argued, the farmer has a legitimate moral right to engage in animal husbandry. The reason for this is that he would be worse-off if he did not raise them. We, as consumers, might only be marginally worse-off if we should forgo meat, but the farmer might face financial ruin. Therefore the liberty principle itself permits him to continue raising animals for food, since not to do so would make him worse-off than the animals he raises.

The problem with this argument, however, is that it neglects a crucial qualification on the liberty principle: *provided that all those involved are treated with respect*. But, as Regan points out, this is a requirement that the present practice of animal husbandry fails to meet. The reason is that present animal husbandry practices treat individuals with inherent value as if they were *renewable resources*. An individual with inherent value is treated as if it were a renewable resource if, before it has reached a state or condition where terminating its life can be defended on grounds of preference-respecting or paternalistic euthanasia, it is killed, its place to be filled by another similar individual whose life will be similarly terminated. Such a practice is unjust because it violates the respect principle. The practice treats an individual with inherent value as if it lacks any independent value of its own, and has value only in relation to the interest of another, that is, the farmer or consumer. A practice of this type treats inherently valuable individuals as *renewable* because it regards them as replaceable without any *prima facie* wrong having been done to those who are killed; and it regards them as *resources* because what value they are assumed to have is viewed as being solely a function of their utility relative to the interest of others. And to treat an individual with inherent value as if it were a renewable resource is, on Regan's view, even worse than treating

them as receptacles. At least when treated as a receptacle, the individual's *goods* (e.g., pleasures or preference-satisfactions) and their *harms* (e.g., pains or preference-frustrations) are viewed as being directly morally relevant to the determination of what ought to be done. When individuals are viewed as renewable resources, however, their goods and their harms have no direct moral significance whatsoever. To view individuals with inherent value as renewable resources is, thus, to view them as even less than receptacles. And, therefore, any practice, institution, or undertaking that permits or requires treating individuals with inherent value as if they were renewable resources, therefore, permits or requires treatment of these individuals that violates the respect principle. As such, the practice, institution or undertaking are unjust. To treat farm animals as renewable resources, therefore, is to fail to treat them with the respect they deserve as individuals with inherent value. But this means that the liberty principle fails to apply, and the farmer cannot, therefore, use it to justify his activities. He could invoke the liberty principle to this end only if those who are harmed by what he does (i.e., the animals he raises) are treated with respect. But, they are not treated with the respect they are owed as a matter of justice, and they cannot be so treated while they are viewed as renewable resources.

As mentioned earlier, the foregoing does not in any way provide an exhaustive account of Regan's treatment of the vegetarianism issue. Nonetheless, it does show pretty clearly that the theoretical framework constructed by Regan provides a clear and consistent account of our duties to non-human mammals. When properly understood, the principles – respect, liberty, miniride, worse-off, and liberty – derived by Regan are more than capable of handling the usual objections to vegetarianism. It is also worth noting that the principles show not only that vegetarianism is morally obligatory. They also show, in an equally clear and consistent manner, why trapping, hunting, and blood sports in general are morally wrong; and why the use of animals in science is wrong. In short, Regan has provided us with a moral theory that is systematic, coherent,

consistent, and of adequate scope to account for our moral dealings with non-human animals. He has, that is, provided us with an extremely powerful moral theory. Nevertheless, it is a theory which I think, ultimately, we should not accept. The remaining sections of this chapter will try to explain why.

10 Inherent value is mysterious

It is possible, I think, to attack Regan both from the point of view of the *content* of the moral principles he adduces and from the point of view of the logical *basis* from which they are adduced. As an example of the former approach, consider Regan's *worse-off* principle. According to this principle, when we must decide whether to override the rights of the innocent few or the innocent many, and when the few would be made worse-off should we adopt the former course than the many would be if we adopted the latter, then we should choose to override the rights of the many. When the harms are not comparable, in other words, numbers don't count. This seems to have counterintuitive consequences. If, for example, we assume that losing one's legs is a harm that is not comparable with losing one's life, that in the latter situation one is made worse-off than in the former, and if we are somehow forced to choose between saving the life of one person and saving the legs of a million others, then the worse-off principle claims that we should choose to save the life of the one. We should allow the million to lose their legs in order to save the life of one person. And this does seem rather counterintuitive.¹² Nevertheless, the criticism of Regan I want to develop in the following pages focuses not on the content of his central moral principles but, rather, on their basis. The basis of these principles is provided by the concepts of *inherent value* and *subject-of-a-life*.

The notion of inherent value is, upon reflection, a mysterious one. What sort of thing is inherent value? Regan is, to say the least, not very explicit about this. One thing, however, is clear. According

to Regan, inherent value exists logically independently of valuers; that is, of individuals who recognize value. That is, for Regan, an individual having inherent value should be clearly distinguished from that individual being valued by others. Being valued by others is certainly not a necessary condition, and probably not a sufficient condition, for an individual having inherent value. It is not a necessary condition because an individual can have inherent value even if it is not valued in this way by others. A subject-of-a-life, such as, for example, a veal calf, raised in isolation from members of its own kind, has inherent value for Regan even though it is not inherently valued by the humans who raise it. For these it is only *instrumentally* valued – that is, valued for its utility relative to the projects of humans. The calf has inherent value even though it is not inherently valued. And being inherently valued may well not be a sufficient condition of having inherent value. We humans, as we all know, have a great capacity to value useless things. This claim does, in fact, raise other issues which we neither need nor want to get immersed in here. It is enough for present purposes to note that Regan wants to clearly distinguish the property of having inherent value from the property of being inherently valued, and all he needs for this claim is that the latter is not a necessary condition of the former. The two properties, therefore, are, for Regan, distinct.

Once we distinguish the property of having inherent value from that of being inherently valued, however, a puzzle emerges. Is inherent value a basic feature of the world, or does it somehow *emerge* from more basic features? Sooner or later, I suppose, scientists might arrive at a complete description of the basic furniture of the universe. And the sort of things which might be referred to in this description are features such as *charm*, *charge*, *quark*, *spin*, and so on. These are basic in the sense that all other things which exist are made up of these sorts of thing. Now, what of inherent value? Is it basic, having the same ontological status as the above items. Or does inherent value somehow *emerge from* or, as philosophers now put it, *supervene upon*, these more basic items. On the one hand, to claim that inherent value is part of the basic furniture

of the universe seems implausible; investigation of the sub-atomic world, it seems, is going to tell us very little about inherent value. But, on the other hand, if inherent value emerges from the combination of more basic items, it seems we need some account of how this is so, under what conditions it arises, and so on.

These are, of course, difficult issues, and it would be unreasonable to expect Regan's theory to answer everything relevant to the moral domain. Inherent value, according to Regan, is a theoretical posit: we may not know exactly what it is, but, if Regan is right, we have to suppose it exists in order to make sense of our reflective moral intuitions (i.e., our considered moral beliefs). This reply is reasonable, but only to a point. The fundamental problem is that Regan has done nothing to clarify the nature of inherent value, and this undercuts the validity of his claim that inherent value is a genuine theoretical posit.

As was explained earlier, the basic structure of Regan's argument for inherent value is what is known as *inference to the best explanation*. And this, in itself, is a perfectly legitimate form of argument, even though it is not deductively valid. However, once we have posited an entity using this argument form, it is then surely incumbent upon us to try and say something about the nature of this entity. More generally, any theorist who wants to posit the existence of a certain theoretical entity, in order to explain a certain range of phenomena, is then placed under a fairly pressing methodological burden to investigate, and to try to say something about, the nature of the postulated entity. Otherwise she has not provided us with a theory at all, but only with a theoretical hole waiting for a theory to be put in its place.

Consider an analogous case. Cartesian dualists claim that the mind is a non-physical entity. One of the arguments for this claim is an inference to the best explanation. No physical system, like the brain, dualists claim, could possibly explain the capacity of people to reason, or use language, or be aware, and so on (the list can be extended, and which feature is focused upon varies from one dualist to another). Therefore, we must suppose that there is

some non-physical part of us which is responsible for our being able to do these things. Now this inference, in itself, is perfectly legitimate – at least, if the premise is true. But, having made this inference, it is then surely incumbent on the dualist to, firstly, investigate the nature of this non-physical mind, and, secondly, show, from this investigation, how this non-physical mind allows us to do the things it was introduced to explain (i.e., to reason, use language, etc.). Dualists have, typically, not even attempted this task. And, therefore, in the absence of this sort of investigation, dualism is seen to be not a theory of the mind at all. And the inference to it was, therefore, not, ultimately, a legitimate theoretical inference either. Dualism, in other words, is not a theory of the mind, it is a theoretical hole or vacuum just waiting for a genuine theory to be put in its place.

The same point can, I think, legitimately be made with respect to Regan's postulation of inherent value. Above all, we must be careful to distinguish two things; two things constantly confused in just about every empirical and quasi-empirical endeavour. On the one hand, there is a theory which seeks to explain a certain phenomenon; on the other, there is an admission that you have no idea what the explanation of that phenomenon is. And, without a serious and concerted attempt to say something about the theoretical entities one has postulated to do the explanatory work, we don't really have a theoretical posit at all. What we have is a simple, though disguised, admission that we have no idea what the explanation is. And, ultimately, I think this is all Regan's postulation of inherent value turns out to be.

11 Inherent value is *ad hoc*

Not only is inherent value mysterious, its postulation is arguably *ad hoc*. To see this, consider the sorts of reason one has for introducing the concept of inherent value, and for postulating it in particular cases. One context in which such postulation occurs is in

connection with the failure of utilitarian accounts of justice. One central problem with utilitarianism, as we have seen, is that, at least *prima facie*, it seems to license all sorts of intuitively unjust practices. That is, there is legitimate moral reason to sacrifice individuals as long as this contributes to the greater overall good (understood in terms of pleasures, preference-satisfactions, or whatever). In order to avoid the counterintuitive consequences of utilitarianism in this regard we must, Regan claims, suppose that certain individuals have inherent value and, therefore, cannot be sacrificed in this way.

Consider, again, the institution of gladiatorial combat discussed in the previous chapter. Suppose society, increasingly inured to violence, decided that traditional sporting fare had become just too tame, and, therefore, decided to reinstate the Roman institution of gladiatorial combat to the death. The gladiators, let us suppose, were selected from people who fulfilled no useful purpose in society, and who had no friends or family. In addition, let us suppose, these people did not want to become gladiators, but were forced to by the authorities on pain of torture and death. Now, it could turn out that the very real misery of the unwilling gladiators was outweighed by the pleasure of the spectators. The extreme misery of a few hundred gladiators, for example, might be outweighed by the individually lesser, but aggregatively greater, pleasure of the several hundred million spectators. If this turned out to be the case, it could be argued, utilitarianism would have to say that instituting this practice of forced gladiatorial combat to the death was the morally right thing to do. But, the argument continues, it is, intuitively, the morally wrong thing to do; it is unjust because it overrides the rights of the victims. To avoid this sort of conclusion, Regan would argue, we must introduce the concept of inherent value. The people who are forced to become gladiators, we must suppose, being subjects-of-a-life, have inherent value. And to force them to fight to the death in this way is to fail to treat them with the respect they are owed as a matter of justice; it is to fail to respect their inherent value. Therefore, we should not, as a matter of justice, treat them in

this way – irrespective of whatever pleasure this might bring to the majority, and irrespective of whatever socially useful functions this practice might have.

Now I'm not suggesting that utilitarianism has no means of circumventing this scenario. There are, in fact, several ways in which, I think, a utilitarian might avoid the conclusion we are asked to draw here. What I want to focus on, however, is one particular way; a way which seems, to me at least, irredeemably *ad hoc*.

Suppose the utilitarian tried to avoid the conclusion of the above scenario by adopting the following strategy. Firstly, the notion of incommensurable benefits and harms is introduced. The basic idea is that certain sorts of benefits and harms simply cannot be compared with certain other sorts of benefits and harms. The two do not occupy the same scale of measurement. Now, if, for example, a certain type of pleasure P_1 is incommensurable with a certain other type of pleasure P_2 , then the two cannot be compared. And it simply makes no sense to ask, for example, how many people have to enjoy pleasure P_2 in order to add up to one person's enjoyment of P_1 . The two pleasures simply cannot be compared in this way. A similar point holds for the attempt to compare incommensurable pleasures and pains. Thus, if a pleasure P_1 is incommensurable with a pain N_1 , then it makes no sense to ask, for example, how many people have to enjoy pleasure P_1 in order to make up for one person suffering N_1 . Again, a comparison of this sort would only make sense if P_1 and N_1 were commensurable. But, let us suppose, they are not. Given this distinction, the utilitarian would then be able to respond to the case of the unwilling gladiators in the following way. The misery suffered by each individual gladiator through being forced to participate in this practice is incommensurable with the pleasure of the spectators. They do not exist on the same scale of measurement. Therefore, it is nonsense to ask how much pleasure has to be enjoyed by how many spectators in order to make up for the misery of the gladiator. No amount of pleasure enjoyed by the spectators could possibly make up for, or balance out, the misery of the gladiator because these particular examples

of benefits and harms are incommensurable; they cannot be compared in the relevant way.

The same approach might then be employed more generally by the utilitarian. That is, whenever we are presented with a case which seems to show that utilitarianism clashes with our considered convictions about justice, we can simply introduce the distinction between commensurable and incommensurable benefits and harms, and assert that, in this case, we are, in fact, dealing with incommensurable benefits and harms; and, therefore, the threat to our considered convictions is only an apparent one.

Now there is a clear sense in which this is an *ad hoc* manoeuvre. On the face of it at least, it is a blatant attempt to save utilitarianism from falsification. Crucially, the method of introducing the distinction is theoretically unprincipled. We bring in the distinction when, and only when, the theory seems to face a serious counterexample. That is, the criteria for when the distinction is to be employed – its criteria of application, if you like – are nothing more than the theory being under threat. And if the distinction is to be both legitimate and legitimately employed, surely some criteria other than the potential falsification of the theory is required. Therefore, the manoeuvre seems to be clearly *ad hoc*.

However, and this is the crucial point, the above manoeuvre of the utilitarian seems to be no more *ad hoc* than the corresponding strategy of Regan. The criterion of application for Regan's notion of inherent value is essentially identical with that of the utilitarian. Regan introduces inherent value when, *and only when*, it is needed to make his theory consistent with his considered moral beliefs. That is the sole criterion for applying the concept in any particular case. Therefore, if the utilitarian's employment of the concept of incommensurable value is *ad hoc* – and it certainly seems to be – there is no reason why Regan's employment of the concept of inherent value is not equally *ad hoc*. It is true that Regan postulates the existence of an entity – inherent value – while our imagined utilitarian posits the existence of a relation – a relation between certain types of harm and benefit. But, whether or not a posit is *ad*

hoc turns only on the criteria according to which the posit is made. The status of the posit itself, and, in particular whether one posits an entity or relation, is irrelevant to this issue.

Therefore, with regard to inherent value, in addition to the postulated entity being essentially mysterious, the postulation itself seems to be seriously *ad hoc*.

12 Inherent value is unnecessary

Postulation of inherent value is also unnecessary. There are two, importantly different, ways of understanding and defending the notion of a moral right.

The first of these is adopted by Regan. We introduce the concept of inherent value; argue that certain individuals satisfy the conditions for possessing this value; and then argue, on the basis of this, that those individuals possess certain moral rights – the right to treatment in accordance with this value, and various other rights derivable from this one. According to this strategy, then, the possession of moral rights is grounded in inherent value, where this is understood as an objective feature of the world.

The second way of understanding the notion of a moral right is based not on the concept of inherent value, but on the concept of *being inherently valued*. That is, the role played by the concept of inherent value in Regan's understanding of a moral right is, in this case, taken over by the concept of being inherently valued. We can define the notion of something being inherently valued as follows:

A thing X is inherently valued by individual I if and only if (a) X is valued by I, and (b) X is not instrumentally valued by I, and (c) X is not subjectively valued by I.

The notion of something being instrumentally valued is explained as follows. You value something instrumentally if you value it

only because of the benefit it affords you. We value money, to use a standard example, only because of what it can do for us, not for what it is in itself. Similarly, medicine is only instrumentally valued by us. Perhaps most of the things we value, we do so only instrumentally. More generally, something is instrumentally valued if it is valued only for its usefulness, its capacity to help us get something else that we want.

Something is subjectively valuable if it is valued only by people who happen to desire it. Lying in the sun drinking margaritas is an activity valued only by people, like myself, and Jimmy Buffett, who happen to enjoy it. As the name implies, something can be subjectively valued by one person and detested by another.

In addition to being instrumentally valued and subjectively valued, however, we have to allow that at least some things are inherently valued. Most of us, in our behaviour, treat some objects as if they are inherently valuable. Thus, they are inherently valued by us. Certain paintings, for example, are inherently valued by some people. That is, many people feel that certain works of art should be respected and protected because of their inherent quality (as they see it) and not simply because people happen to enjoy looking at them (although this is also a legitimate reason). For many people, the theft and possible destruction of Edward Munch's *The Scream* was a disaster, and its return a relief, even if they themselves did not particularly like the painting. The work was inherently valued by those people.

Similarly, one is supposed to inherently value one's children. That is, you are not supposed to value your children, if you have them, because you find their presence entertaining or otherwise pleasurable (subjective value). Nor are you supposed to value them because they provide a useful insurance policy for later in life – someone to look after you in later life (instrumental value). Rather, you are supposed to recognize that your children are valuable individuals in their own right: that they have a value that does not reduce to subjective or instrumental value. To understand your children in this way is to value them inherently.

It seems fairly clear, then, that at least some things are inherently valued by people. It is important to realize that the fact that certain things are inherently valued does not entail that they are inherently valuable. People who find things inherently valuable could simply be deluded. There could be no such thing as inherent value, it could be a metaphysical illusion, but still be the case that people treated certain objects *as if* they were inherently valuable. That some things are inherently valued, then, entails that people treat them as if they were inherently valuable. It does not entail that they actually are inherently valuable. And to claim that some things are inherently valued by some people, or all people for that matter, does not commit one to the claim that those things are inherently valuable. The claim that some things are inherently valued, then, commits one to a lot less than the claim that those things are inherently valuable.

The reason why Regan's postulation of inherent value is unnecessary is because precisely the same role played by that concept in deriving a rights-based moral philosophy can also be played, instead, by the concept of being inherently valued. That is, beginning with the concept of certain things being inherently valued, we can erect substantially the same edifice of rights and obligations that Regan builds on the bedrock of inherent value. Thus, for example, whereas Regan, starting from inherent value, derives the duty to treat things with such value in ways which respect this value, we, starting from the concept of being inherently valued, can derive the duty of treating things which are inherently valued in ways which respect the fact that they are inherently valued. And, whereas Regan, starting from inherent value, derives the *prima facie* duty not to harm individuals with this value, we, starting from an individual's being inherently valued, can derive the *prima facie* duty not to harm individuals who are valued in this way. In fact, a recognizable form of Regan's rights-based theory can be constructed on the basis not of inherent value, but on the less ontologically dubious fact that certain things are inherently valued. This is, fundamentally, the project of Chapter 6.

But surely this project is doomed to fail? Surely the project of deriving a rights-based view from the fact that certain things are inherently valued is certain to founder on the fact that different things are inherently valued by different people? And surely it will founder on the fact that many people, most people in fact, do not inherently value non-human animals? Is this not more than evident in their treatment of them? And this, of course, is all true. And while this is not the place to anticipate the arguments of Chapter 6, I can say this. In the project of deriving a rights-based view from the concept of being inherently valued, we have more to work with than just the claim that certain things are inherently valued by certain people and that certain things are not. That is, we have more to go on than just the brute facts about what people, as a matter of fact, inherently value (i.e., profess to find inherently valuable). That is, our focus is not restricted to individuals and their behaviour, but also encompasses the principles that these individuals have adopted in order to regulate their social interactions. Thus, what we also have to go on is the fact that people, whatever society they may belong to, belong to a social structure that will be built on fundamental principles; principles which regulate what things people should find, and treat as if they were, inherently valuable. And these fundamental principles, whether we realize it or not, entail that we should find other things inherently valuable also. Thus, we are committed by the fundamental principles upon which our society is based, to treat certain things as if they were inherently valuable. And this is true independently of whether we do in fact treat them in this way. We are committed, then, on pain of inconsistency, to treat certain things as if they were inherently valuable – whether we know it or not, and whether we do in fact so treat them or not. And, I shall argue, we, as members of a society built broadly on the principle of equal consideration – and the underlying idea that there can be no moral difference without some relevant other difference – are committed to treating non-human animals as if they were inherently valuable; whether we know it or not, and whether we do so or not.

If this is correct, then a rights-based view capable of underwriting the claim that human animals have substantial direct moral duties to non-human animals does not need to be based on any assumption of inherent value. And this is a good thing. As Rawls has pointed out, the fewer controversial assumptions a moral theory is based on, the wider its force and appeal is likely to be. The assumption that certain things have a mysterious property of inherent value is a controversial metaphysical assumption if ever there was one. Therefore, all things being equal, it is better to avoid basing one's moral theory on it if at all possible. The arguments of Chapter 6 will try to demonstrate that it is possible.

First, however, I want to look at whether a serious case for the moral standing of animals can be made on the basis of what is known as *virtue ethics*.

5

Virtue Ethics and Animals

1 The concept of virtue?

The expression ‘virtue ethics’ denotes a relatively loose tradition of ethical thinking that, in the West, stems from Aristotle (and to a lesser extent Plato) and, in the East, has identifiable roots in Chinese philosophy, particularly Confucianism. In more recent Western philosophy, virtue ethics became largely moribund after the Enlightenment (although was certainly the dominant approach before that), and stayed that way until it was revived, almost single-handedly, by Elizabeth Anscombe’s article ‘Modern moral philosophy’.¹

As the name indicates, the central concept of virtue ethics is that of a *virtue*. A virtue is a *character trait* that is deeply entrenched in its possessor and also, crucially, multi-factorial. To say that it is deeply entrenched in its possessor is to say that it manifests itself in more than a single type of action, and this manifestation will be stable through time. Thus, the virtue of honesty will manifest itself not just in the fact that I do not steal from others, but also in the fact that I will do my best to return what others have lost (rather than pocketing it for myself). And these sorts of behaviours are not ones I exhibit sporadically, but are relatively constant through time. All things being equal, I will return the lost money not merely today, but on any day that I find it. To say that a virtue is multi-factorial is to say that it consists in more than behavioural tendencies or dispositions

alone, even if these are stable through time. To have the virtue of honesty, for example, is not just the tendency to do honest things. It is also the tendency to deplore dishonesty in oneself and others, to feel outrage when one witnesses this dishonesty, and to make this outrage known; and so on. In order to be constitutive of a virtue, the stable behavioural dispositions must be located in an appropriate surrounding context of judgments and emotions. Bound up in the possession of a virtue, therefore, is not a single disposition to do certain things in given circumstances, but also the disposition to have judgments, emotions, thoughts, feelings, and so on that are 'appropriate' to these circumstances. The reason for this is pretty clear. A person can have the (stable) tendency to do what is honest and refrain from doing what is dishonest because, and only because, she is afraid of being caught, and of the sanctions that would inevitably result. Since, in this case, the tendency to do what is honest and refrain from doing what is dishonest is not situated in the appropriate surrounding milieu of emotions, judgments, and other evaluative acts, her tendency is not part of a virtue of honesty. She possesses no such virtue. Therefore, it would be unwise to attribute a virtue to a person on the basis of observing their actions – even if these actions are consistent through time – if one does not know the reasons for actions. In the possession of a virtue, actions, judgments, and emotions are bound up in an indissoluble whole.

These considerations lead us in the direction of a definition of the concept of a virtue that, while not unassailable, is good enough for the purposes of this chapter. Roughly speaking:

A virtue is: (i) a good, admirable, or otherwise praiseworthy *character trait*, where (ii) a character trait consists in a relatively stable set of behavioural dispositions that are embedded in an appropriate surrounding milieu of judgments and emotions (broadly understood).

The corresponding notion of a vice can then be defined as a bad, unworthy, or blameworthy character trait, where we understand

the notion of a character trait in the same way. The concept of virtue is, of course, correlative to the concept of vice. To have a virtue is, at the same time, to abhor the corresponding vice. Part of the attraction of virtue ethics lies in the complex vocabulary of virtues and vices we possess in ordinary language. The vocabulary of the virtues is both large and rich: 'courage', 'kindness', 'honesty', 'justice', 'benevolence', 'loyalty', 'industriousness', 'altruism', 'generosity', 'compassion', 'responsibility', 'mercy', 'integrity', 'wisdom', and so on. The vocabulary of the vices is, if anything, larger, richer, and more detailed. We condemn people for, among other things, being: 'cowardly', 'cruel', 'dishonest', 'unfair', 'malevolent', 'disloyal', 'lazy', 'self-serving', 'mean', 'callous', 'irresponsible', 'foolish', 'greedy', 'spiteful', 'jealous', 'envious', 'arrogant', 'small-minded', 'uncaring', and so on. Neither of these lists is exhaustive. But they do indicate the extent to which our everyday language is a rich evaluative depository of virtue and vice terms. This vocabulary provides the basis of the virtue-ethical approach toward moral issues, problems, and disputes.

Armed with the concept of a virtue, we can then define the virtuous person as one who has, and exercises, the various virtues – understood as entrenched, multi-factorial character traits in the sense outline above. Since having and exercising a given virtue precludes having and exercising the corresponding vice, a virtuous person is one who acts according to virtue (and so does not act according to vice). A virtuous person, in short, is one who acts virtuously. According to virtue ethics, the fundamental moral injunction is for one to be, or become, a virtuous person – and for anyone else to do likewise.

2 Virtue ethics and animals: Scruton versus Hursthouse

It is common to find three objections levelled against virtue ethics (perhaps not always properly distinguished by those who level them). The first is the charge of *subjectivity*: one's person virtue is another person's vice, and conversely. One person may regard honesty as a

virtue; another may regard it as a sign of naivety or weakness. Who is to say who is right? The second is the charge of *vagueness*. This is distinct from the charge of subjectivity. The objection is not that the application of virtue terms is a matter of subjective feeling or personal opinion. Rather the charge is that there is enough leeway in their application to leave plenty of opportunities for dispute. The third charge is that virtue ethics is committed to the possibility of crippling *conflicts* of virtues. For some activities at least, exhibiting a given virtue is possible only if we also exhibit a particular vice. What can virtue ethics say in the case of this sort of conflict?

In this section, I shall not make much of the charge of subjectivity. If you are a subjectivist about the virtues and vices, then this is probably because you are a subjectivist about morality as such. If so, then your concerns are almost certainly outside the scope of this book. While not necessarily endorsing them, the charges of *vagueness* and *conflict* provide a useful means of orienting oneself in the discussion to follow. Both the strengths and weaknesses of virtue ethics are, I think, best revealed in its discussion of particular ethical issues or dilemmas. Therefore, I propose to jump straight in and discuss the case of animals. I am going to examine the dispute between Rosalind Hursthouse and Roger Scruton over the moral status of blood sports. The dispute is interesting; both protagonists adhere to a version of virtue ethics, but reach very different conclusions. The question we need to look at is whether this reveals a glaring deficiency in the theoretical apparatus of virtue ethics, or whether, on the contrary, this apparatus supplies us with a way of adjudicating between the two.

In employing what is essentially a virtue ethical approach, Scruton argues in favour of practices such as fox-hunting (and the eating of animals for food).² Hursthouse, on the other hand, argues that virtue ethics precludes such practices.³ Scruton argues that whether practices such as fox-hunting, bullfighting, angling, and so on count as morally legitimate is crucially dependent on the motives of the persons engaged in them. If, for example, a person were to derive a sadistic pleasure in watching a fox being torn apart

by hounds, and seeing a bull suffer a slow and painful death from exhaustion and blood loss, then in engaging in, or being a spectator of, these activities, that person would not be acting in a virtuous way. They would be guilty of the vice of cruelty. Scruton, however, claims that cruelty, or as he puts it, 'sadism towards the fox', is rarely one of the vices displayed by participants in these activities. Rather, the suffering of the fox or bull, is generally regarded as an unfortunate but inevitable consequence, or by-product, of an activity that, when properly conducted, can involve the exercise of important virtues. Fox-hunting, according to Scruton, counts as the deliberate embarking of an action of which pain is an inevitable but nonetheless, *unwanted* by-product. As long as the participants in the fox hunt do not enjoy the suffering of the fox, or pursue their activity with the motive of making the fox suffer, then, Scruton argues, fox-hunting does not involve the vice of cruelty. Here, Scruton is, in effect, utilizing what we identified earlier as the holistic or multi-factorial nature of virtues: the fact that in any exercise of a virtue, action, judgment, and emotion are indissolubly connected.

Hursthouse disagrees with Scruton's analysis. It may seem as if Scruton is simply making an empirical claim about the motives of fox-hunters – one which stands or falls on the results of a proper investigation of their motives (whatever that might be). But the real problem, Hursthouse argues, is that Scruton is unrealistically restricting his use of the term 'cruelty'.⁴ We can agree, as all virtue ethicists must, that the deliberate infliction of pain in order to achieve some other purpose to which that pain is a necessary means, is not necessarily cruel or callous. I may need to have my dog subjected to a painful medical procedure in order to save its life. Nevertheless, in the case of fox-hunting, bullfighting, and other blood sports, the participants and/or spectators are at least guilty of the vice of callousness. The participants or spectators are fully aware of the pain and suffering inflicted on the animals; it is just that this pain and suffering does not matter to them as much as it should. They are acting callously, and this is not what a virtuous person does.

Recall now the three charges often levelled against virtue ethics that it is: (i) *subjective* in its application of virtue terms, or (ii) unacceptably *vague* in this application, and (iii) subject to crippling *conflicts* of virtues and vices in particular cases. If these charges are correct, we will not, I think, find them here – at least not yet. First of all, with regard to the issue of subjectivity, what we seem to find, so far, is a mistake – of a perfectly objective nature – on the part of Scruton. Scruton has been at pains to deflect a certain sort of criticism of blood sports: that they exhibit the vice of cruelty. But if he has successfully deflected this criticism, it is only because he has completely overlooked another vice – callousness. Even if (and, please note, I say *if*) the participants in or spectators of blood sports are not guilty of cruelty, they are certainly guilty of the vice of callousness. There is nothing subjective about this mistake, and therefore nothing to justify the charge of subjectivity.

Consider, now, the issue of vagueness. There is some evidence that the dispute between Scruton and Hursthouse is abetted by a certain amount of vagueness in the use of the term ‘cruelty’. Cruelty, for Scruton, seems to involve the deliberate infliction of pain or suffering for its own sake: that is, because the infliction of pain and suffering is desired for its own sake by the one who inflicts. One might legitimately question this conception of cruelty – although it is, I think, by no means idiosyncratic to think of it in this way. In her reply to Scruton, however, Hursthouse can, at least in some places, be accused of running together the concepts of cruelty and callousness in a way that Scruton would not (and should not) accept. For example, she writes: ‘Watching bullfighting is acting cruelly and callously, notwithstanding the putative fact that the spectators take no pleasure in the animals’ suffering’.⁵ It might be callous but, from the perspective of the definition of cruelty presupposed by Scruton, it cannot be cruel.

Nevertheless, while there is evidence of some vagueness in the dispute, there is no reason for thinking that this is in any way fatal to the project of a virtue ethical approach to moral disputes. We have a possible case of vagueness, admittedly, but it is nothing

that cannot be sorted out by a proper – and, crucially, agreed upon analysis of the concepts of the virtues and vices. Providing such an analysis is, of course, one of the central tasks of virtue ethics. For our purposes, however, I propose to resolve this type of disagreement between Scruton and Hursthouse by, in effect, giving Scruton cruelty and giving Hursthouse callousness. Scruton, we can accept, has shown that blood sports are not necessarily cruel. But they are nonetheless, as Hursthouse has convincingly shown, callous.

Since a virtuous person does not act according to vice, and since callousness is undeniably a vice, then it might seem that we have the basis of a virtue ethical case against blood sports. This conclusion would, however, be premature. We still have the possible problem of *conflict*. The first thing we have to do is be very careful in identifying the nature of this conflict.

To begin with, here is an objection that would not, in any way, inconvenience Hursthouse. We might point to the fact that fox-hunting, for example, brings with it other benefits. Fundamentally, it is *fun* – or at least so its proponents claim. Scruton seems to agree with this – indeed, he seems to think that is its fundamental purpose. So too is bullfighting, and angling, and other blood sports. So, one might object that while such activities do indeed exhibit the vice of callousness, this is overridden by the fun that they also occasion. This sort of objection is not the sort of thing that can inconvenience Hursthouse or, indeed, any virtue ethicist. A fundamental tenet of virtue ethics, of course, is that when there is a conflict between virtue and fun, virtue is going to win. Virtues, much like rights for Regan, are moral *trumps*. From the perspective of virtue ethics, virtues trump things that are not virtues. Intuitively, this does seem right. Consider what we would say of a person who tortures animals, not in the context of a blood sport, but in the privacy of his own home. He does it because he finds this fun. Being fun is undoubtedly a benefit, but not the sort of benefit that could outweigh the vice of cruelty (and so outweigh the person's failure to be virtuous).

The problem of conflict for virtue ethics arises not when virtues conflict with other things that might be thought beneficial, but when virtues conflict with other virtues. Thus, Scruton is on much stronger ground when he argues that there are certain virtues associated with blood sports such as bullfighting or fox-hunting. In particular, Scruton claims, participants in these sports must exhibit the virtue of *courage*. Matadors are killed in the pursuit of their sport. And occasionally fox-hunters are thrown from their horses and break their necks. To engage, voluntarily, in a dangerous pursuit, Scruton argues, is to exhibit the virtue of courage. Therefore, while fox-hunting and other blood sports exhibit the vice of callousness, they can also exhibit the virtue of courage. We have a conflict between a vice and a virtue. Which, from the perspective of virtue ethics, is going to win out? Indeed, can virtue ethics, by itself, provide us with the resources to make a decision one way or the other?

There are at least two options available to the virtue ethicists at this point. The *first* is to deny the possibility of such conflict. The apparent conflict between a virtue and a vice in this instance is just that – apparent. There is no real conflict. This option, in turn, can take two forms. The first form appeals to the idea of a fully virtuous person. When a fully virtuous person acts according to the virtues she possesses, it is impossible for her, at the same time, to act according to vice. In the fully virtuous person, acting according to virtue precludes acting according to vice. This version of the option, however, does not concern us. Few of us are fully virtuous persons: I know I'm not. And I'm far from convinced that many – perhaps any – of us have any clear idea of what is and what is not possible in the case of a fully virtuous person. More interesting in the present context is a stronger version of the option which looks like this: if, when engaging in a particular activity, one is acting according to virtue, then it is impossible for one to be simultaneously acting according to vice.

Intuitively, this claim seems to me to be impossibly strong. However, interestingly, Hursthouse seems to pursue a version of

this idea, at least in the case of fox-hunting. Her strategy, therefore, is to deny that participants in fox-hunting exhibit the virtue of courage. To this end, she accuses Scruton of misunderstanding the concept of courage:

...used in the context of virtue ethics, as a virtue term, 'courage' does not mean simply 'facing danger'; it means 'facing danger for a good reason, or for a worthwhile end' and is contrasted with 'daring' and 'reckless', both of which standardly count as vices or faults. If, in some very peculiar circumstances, leaping one's motorbike over cars was going to save someone's life, or avert some disaster, it would be courageous. But if it is only done, as reckless teenagers do it, for fun and excitement and to terrify the people in the car park, it is not courageous and not the sort of thing a virtuous person would do.⁶

So, participants in blood sports are not, in fact, being brave because their actions are not performed for a *good reason* or *worthwhile end*. Here, Hursthouse is locating herself solidly in a tradition of thinking about courage that derives from Aristotle. Apparent acts of courage not performed for a good reason or worthwhile end are, in reality, acts of recklessness rather than courage. Just because an idea has pedigree, of course, does not mean that it is correct. But, nevertheless, for the sake of argument, let us give Hursthouse this premise: in order to qualify as an expression of the virtue of courage, an act must be performed for a good reason or worthwhile end. This, Hursthouse thinks, settles the dispute with Scruton in her favour. Scruton is guilty of misunderstanding the nature of courage; and that is why he thinks fox-hunters can exhibit this virtue:

Now it would be foolish to insist, in the teeth of the dictionary definition, that the word 'courage' *really* means 'facing danger for a good reason, or for a worthwhile end'. But it is reasonable to insist that Scruton, talking in terms of the virtues and vices (and well acquainted with the many philosophical texts in which the restricted meaning is emphasized), should use it in the restricted way, not the standard dictionary one.⁷

Scruton has, she thinks, carelessly relied on the dictionary explanation of courage, and ignored the preferred understanding of this concept embodied in the virtue ethical tradition. If this is Hursthouse's argument, however, then she has seriously misunderstood the nature of the dispute between her and Scruton at this point. The restricted conception of courage as facing danger for a good reason or a worthwhile end is one that is, in fact, readily available to Scruton. He need simply point out that Hursthouse has, in effect, a restricted conception of what constitutes a worthwhile end. This point is masked by Hursthouse, in effect, artificially restricting the possible ways in which something might be a good reason. In the motorbike passage cited earlier, she presents the available options as either saving someone's life or averting some disaster (good reasons) or fun, excitement and terrifying people in the car park (bad reasons). One can accept that the former are good reasons and the latter are bad ones, but still insist that these do not exhaust the possibilities.

Consider some of these other possibilities. Someone might become very nervous before they surf a big break. Or they might be terrified before they make a difficult ascent of an icy cliff face. Or someone might be frightened before the beginning of some contact sport – rugby, boxing, gridiron, and the like. When asked why they do this, they respond something like this: they do it partly because they enjoy it, but also partly because they think facing their fear is a good thing to do: facing fear is good reason, or a worthwhile end, for an activity that produces this fear. If pushed further on why they think facing their fear is a good thing, they might respond in one of two ways. Some might point out that life is scary; the world is an often frightening place to be. And if they are able to face and contain their fears in these sorts of restricted areas, they are more able to bring this ability to bear on other areas of their life. Others might simply respond that facing their fear – even if they are largely responsible for producing it – is an inherently ennobling thing to do; that they feel it makes them stronger or better in some vague and difficult to define way. We don't need to look to the sporting

arena for examples of this sort. Someone might think that one of the benefits of moving to a new and unfamiliar country or city is the fear that the prospective move occasions, and that facing this fear is a good thing to do for one or other of the above reasons.

Should we immediately and unconditionally dismiss these reasons as not good or worthwhile ones? These explanations of why facing fear is a good thing to do may not be to everyone's tastes. But they certainly resonate with many. And it is far from clear that the virtue ethicist's armchair supplies us with the warrant to simply dismiss them. If facing fear in certain restricted arenas does help, as it is sometimes put, 'build character' or if it does help you deal with the slings and arrows that outrageous fortune throws at you more generally – and both of these are empirical claims – then we have no basis for simply dismissing them *a priori*. And if Hursthouse were to dismiss them, there is no reason why Scruton should follow her in this.

On this point, however, Hursthouse might employ the second prong of her attack on Scruton's claims concerning the courage involved in fox-hunting. In her second prong, she accuses Scruton, and participants in blood sports in general, as confusing courage with *vainglory*.

[Scruton's] claim that fox-hunting 'displays and encourages' the virtue of courage, in himself and others, would show, in Mary Midgeley's words, 'the glaring faults of confused vainglory and self-deception' ...I would say that similar remarks apply to Scruton, and indeed, most strikingly, in two of the writers he mentions with approval, Trollope and Surtees. Both of them constantly describe the fox as 'a noble adversary' and the hunt as Homeric, a noble battle, 'a contest with the quarry', encouraging in themselves and their readers the ludicrous fantasy that the contest between one small animal (albeit, as they always say, 'a remarkably cunning one') and a pack of riders, horses and hounds, is comparable to the war between the Greeks and the Trojans.⁸

It is, of course, a ludicrous fantasy, and it is difficult to feel anything but contempt for Trollope, Surtees, and anyone who thinks

of their fox-hunting endeavours in this way. Anyone who thinks in this way does, without a shadow of a doubt, have serious vain-glory issues. But is this the only way of looking at the matter? For example, what if the person were to reflect on their activities in this way? 'OK, I got up on my horse, and I – or rather my horse – ran quite fast, jumped over a few hedges, and so on. I am aware of the fact that people do fall off their horses in these sorts of circumstances. Sometimes they are injured, crippled, or even killed. This is rare, I know; but it does happen. So, I was a little scared when it was happening, but I'm glad I did it for that reason.' There is nothing here suggesting a Homeric war with the fox, and so, it seems, nothing that would merit the charge of vainglory. A person who engages in other dangerous sports might look at things in the same sort of way.

Hursthouse could, of course, contest this. In particular, she might accuse the person in question of over-estimating their bravery. Very occasionally people are killed or seriously injured while fox-hunting. But it doesn't happen very often (most serious injuries involving horses happen when the person is not actually on the horse – kicks administered whilst grooming being the usual culprit). In fact, the chances of death or serious injury are negligible. So, it is overwhelmingly likely that one will emerge from the hunt unscathed. Therefore, the fear one seeks to overcome through completing the hunt is misplaced. If this is correct, then the charge of confusing courage with vainglory has not, in fact, been disarmed: to suppose that one is in terrible danger when in fact one is not, is, itself, a form of vainglory.

This option, however, is unappealing in one crucial respect: it holds people up to a standard of rationality in their risk assessment that few people, if any, are able to meet. More precisely, it requires people to bring their subjective sense of risk and objective assessments of risk into line – a task that is beyond most of us, perhaps all of us. To take one example, consider the 7/7 tube and bus bombings that took place in London, in July 2005. Fifty-two people were killed and a little over 200 were injured. Suppose there was a similar attack every working day. Then roughly 1000 people a week would

be injured. That's roughly 50,000 injuries a year – when allowing for holidays. In 20 years, we have a million injuries. However, approximately eight million people use the London public transport system on any given working day. So, if I travelled to work five days a week, for 20 years, and there was an attack of the 7/7 level of magnitude every single working day for those 20 years, my chances of being injured in those 20 years would still only be one in eight. And that is after 20 years. So, the chances of me being injured by one of these attacks on any given day would be negligibly small. This is a matter of objective risk assessment. Nevertheless, it would be a brave person indeed who was not a little nervous going to work the day after 7/7. Our subjective sense of risk is rarely even on nodding terms with objective levels of risk. And, in many ways, this is a good thing. Objective assessments of risk are all very well, but we only have to be wrong *once* and that's all she wrote. That is why our subjective sense of risk is wildly inflated relative to the actual risks objectively present in the environment. To require that our subjective sense of risk be brought into line with objective assessment of risk is one that is very difficult, and perhaps impossible, for us to achieve. Therefore, it is difficult to see the legitimacy of any demand that we do so. Or, to put the same point another way: it is *natural*, *useful*, and given that one mistake might mean it is all over for us, arguably *rational* to be afraid in situations where the objective risk is very small – even negligibly small. It is not irrational to be scared jumping out of a plane even if one's parachute and instructor are both reliable. The risk is small, but the potential consequences are huge. I think one can say the same thing about other potentially dangerous sports such as big break surfing, rugby, rock-climbing, and also fox-hunting and bullfighting.

Therefore, I do think a case can be made that people who engage in fox-hunting do exhibit the virtue of courage. Hursthouse's contention that they do not rests on an unjustifiably restrictive conception of what constitutes a good reason or worthwhile end for facing danger. I don't think this conclusion is in any way surprising. Unspeakably evil people can be astonishingly brave.

And, crucially, they can be unspeakably evil and astonishingly brave at the same time and in doing the same thing. US comedian Bill Maher was fired for saying this – but he is obviously correct: courage was not one of the deficits of the moral monsters that perpetrated 9/11.

In seeking to deny that people who engage in fox-hunting can exhibit the virtue of courage in the commission of this activity, Hursthouse is leaning towards the idea that one cannot, in the commission of the same activity, exhibit both a virtue and a vice. There is, however, another option available to the virtue ethicist: allow the possibility of such conflicts, but find a way of ranking the various virtues so that we can work out, in any given case, whether the virtue or vice exhibited should be given precedence. A promising way of pursuing this strategy is by distinguishing between the *moral* virtues and the *executive* virtues.⁹ Moral virtues are ones that track moral values. Examples would include kindness, honesty, justice, benevolence, loyalty, altruism, generosity, compassion, responsibility, mercy, and integrity. Executive virtues are ones that do not track moral values, and might include courage, industriousness, and wisdom. Given this distinction, a virtue ethicist might argue that moral virtues always trump executive ones. Consequently, when we have a conflict between an executive virtue like courage, and a vice such as callousness, the exhibiting of callousness should be regarded as a more serious failure than the exhibiting of courage is regarded as a success. The virtuous person will, therefore, seek to eliminate their exercise of callousness, even if in this case it also means eliminating their exercise of courage.

This does seem an intuitively plausible way for virtue ethics to address the problem of conflict – assuming we can make the case that moral virtues always trump executive ones. However, virtue ethics now starts to look oddly familiar. We have already seen how virtues play the role of trumps in moral disputes – in much the same way that, for people like Regan, rights play the role of trumps in moral disputes. It is not possible to justify blood sports by appealing

to the fun they engender, since they also involve the vice of callousness. And not exhibiting this vice – and so, all other things being equal, being a virtuous person is more important than having fun. Similarly, Regan would argue that we cannot justify an activity like fox-hunting by appeal to the fun that it involves, because this fun is purchased at the expense of the rights of the fox, and no amount of fun can justify overriding a right.

The approach to conflicts of virtues based on the distinction between moral and executive virtues further increases the parallels with rights-based approaches. What we now need are priority principles that specify when one virtue trumps another – priority principles of roughly the same kind that Regan was at pains to delineate in developing his rights-based account. The game now starts to take on a distinct air of familiarity: the cards have changed but the game is pretty much the same; a game of trumps, and meta-trumps (and perhaps meta-meta-trumps). One can't help suspect that, for all its claims to offer a genuine alternative, virtue ethics is treading the same region of logical space as its utilitarian and rights-based competitors.

However, we need not become embroiled in these issues. With regard to the issue of fox-hunting, it is clear why Scruton does not have a moral leg to stand on. Let us suppose he is right in claiming that fox-hunting involves the virtue of courage. Hursthouse is also correct in her assessment that it involves the vice of callousness. We do not need to become involved in the issue of which – the virtue or the vice – should take precedence. The most obvious question is: can we replicate the conditions under which the virtue is exercised whilst eliminating the exercise of the vice. And it is pretty clear that we can. That is what drag-hunting is for. More generally, if you, like me – old adrenalin junkie that I am – think that inducing and then facing fear can be an intrinsically good thing to do, then the least you can do is try to do this in situations that don't involve the exercise of any countervailing vice – for that, to say the least, is going to take the moral shine off what you do. So if, as Scruton argues, fox-hunting is about courage,¹⁰ then why not try to find ways

of exercising that courage that does not involve the painful slaughter of defenceless animals. Try jumping out of a plane.

3 The virtue of mercy

In this section, I want to develop some of the points gleaned from the dispute between Hursthouse and Scruton into a more general virtue ethical case for animals. I am going to argue that animal plausible virtue ethics will preclude most of the ways we currently treat non-human animals.

I shall begin the argument with a claim made, with his usual sagacity, by Milan Kundera in *The Unbearable Lightness of Being*:

True human goodness can manifest itself, in all its purity and liberty, only in regard to those who have no power. The true moral test of humanity (the most radical, situated on a level so profound that it escapes our notice) lies in its relations to those who are at its mercy: the animals. And it is here that exists the fundamental failing of man, so fundamental that all others follow from it.¹¹

Here, Kundera identifies what he thinks of as the ‘true moral test’ of humanity, and at the same time identifies a certain virtue that is crucial to this test: *mercy*. This virtue and its corresponding vice – *mercilessness* – are peculiarly salient to our dealings with those who, relative to us, have no power. And, as Kundera notes, animals provide the most obvious examples of those who have no power. I think Kundera is right to allocate mercy this central role amongst the moral virtues. In this section, I shall explain and defend this centrality.

In developing this argument, it is crucial to remember the multi-factorial character of the virtues. Bound up in the possession of a virtue is far more than merely being disposed to behave in certain ways in given circumstances, even if this disposition is stable through time. Virtues are not merely dispositions to behaviour. Rather, any such dispositions must be surrounded by, and

grounded in, a milieu that consists of the relevant judgments and emotions. This claim is essential to any plausible virtue ethics.

With this in mind, I shall argue that mercy is fundamental to the moral virtues in that it is required for – a necessary condition of – many of the other moral virtues. I shall not argue that it is required for possession of all the other moral virtues. I suspect that it is, but this is not required for the argument I am going to develop. To see why, consider someone who fails to exhibit the virtue of mercy. The person is, let us suppose, exemplary in their dealings with those who have power – which we can understand, in a sense that is rough but sufficiently precise for our purposes, as those who are capable of helping or hurting them. However, when they come to interacting with the powerless (i.e., those not capable of helping or hurting them), they fall short of this standard in some or other respect. Development of this argument does not require us to say what it is for them to be exemplary in their dealings with those who have power, nor does it require us to specify the way in which they fall short of this standard in their dealings with those who do not. With this at least rough-and-ready scenario in mind, let us consider some of the more important moral virtues.

The virtue of kindness is an obvious place to start. We are to try to imagine a scenario in which someone exhibits the virtue of kindness towards those who are capable of helping or hurting him, but fails to exhibit this virtue towards those who are not. This, I shall argue, is not a possible scenario. Such a scenario is *apparently conceivable*; but it is not *genuinely possible*. It is apparently conceivable because we can imagine a scenario that seems, to us, to be one in which a person is kind only toward those who have power. But it is not genuinely possible because we have, in fact, succeeded only in imagining something else. What we in fact end up imagining is a scenario in which the person's behaviour towards those who have power bears all the hallmarks of behaviour that we would call kind. However, this is not, as we have seen, sufficient for the possession of the virtue of kindness. For sufficiency, we need to supply the surrounding context of emotions

and judgments. The question is, then, can this context plausibly be supplied in the scenario we are trying to imagine? It is difficult to see how it could. For the person's failure to behave in a similarly kindly way to those who do not have power – for his behaviour to fall short of whatever standard he achieves with respect to those that do have power – seems inevitably to indicate that his, as we would put it, 'kindly' behaviour towards those who have power is motivated by something other than kindness. That is, it is motivated by something other than the sort of judgments and emotions that partly constitute the virtue of kindness. The motivation seems coloured by considerations of self-interest – for what else would explain the difference in his behaviour towards those who have power and those who do not? However, if the surrounding judgments and emotions are not in place, then the person's 'kindly' behaviour towards those who have power is not in fact a manifestation of the virtue of kindness. So, the situation in which a person exhibits the virtue of kindness in the absence of the virtue of mercy is not, in fact, a possible situation. It might be apparently conceivable; but it is not genuinely possible. If this is correct then possession of the virtue of mercy is a necessary condition of the possession of the virtue of kindness.

The same sort of argument can be applied to cognate or closely related moral virtues such as compassion, generosity, and benevolence. If one's 'generosity' extended only as far as those who were able to help you or hurt you, and was markedly curtailed in the case of those who were not, then the conclusion we should draw is that this is not a 'genuine' case of generosity. That is, the behaviour is not an exemplification of the virtue of generosity. It is not a genuine case of generosity because the surrounding judgments and emotions that would make it so are not in place. So, once again, we might think that we can imagine someone who is generous only in her dealings with those in a position to help or hurt her, but falls short of this in her dealings with those who are not capable of these things, but what we think we can imagine is not a possible situation. Neither can we, for essentially the same reasons, really

succeed in imagining someone who is benevolent or compassionate only in his dealings with those who have power.

Consider, now, another important moral virtue: *loyalty*. Can we really imagine someone who is loyal only towards those who are in a position to help or hurt him, and falls short of this in his dealings with those who are not? Once again, this does not seem to be a genuine case of loyalty. The obvious question is: what would happen if those who are in a position to help or hurt him suddenly, perhaps through some or other misadventure, lose this ability? In the scenario we are trying to imagine, the person would then, in his dealings with these people in their newly diminished circumstances, fall short of the loyalty he previously seemed to exhibit. If this were so, then we should deny that the behaviour he previously exhibited was a manifestation of the virtue of loyalty. The reason is that the surrounding context of judgments and emotions was not in place, and without this the person's behaviour, while ostensibly loyal, was not, in fact, loyal at all. That is, it was not an expression of the virtue of loyalty. One cannot possess the virtue of loyalty if one's seemingly loyal behaviour is restricted to those who have power. And this is equivalent to saying that the virtue of mercy is a necessary condition of the virtue of loyalty.

A similar argument applies, without significant revision, to the virtue of honesty. Someone who is honest only in her dealings with those who have power, but falls short of this standard when dealing with those who do not, is not, we can legitimately say, 'really' honest. Their seemingly honest behaviour is not situated in a surrounding context of emotions and judgments required for it to be an expression of the virtue of honesty. We might think we can imagine someone whose honesty is restricted in this way. But what we are not thereby imagining is a case where the virtue of honesty is restricted in this way. We are imagining a certain sort of behaviour, admittedly; and this behaviour might certainly seem to be a case of honest behaviour. But it is not, in fact, a manifestation of the virtue of honesty. The virtue of mercy is a necessary condition of the virtue of honesty. The same sort of argument applies,

again without significant revision, to cognate moral virtues such as integrity.

4 Conclusion

The virtue ethical defence of animals turns on acknowledging the peculiar centrality of the virtue of mercy. The virtue of mercy is a peculiarly foundational moral virtue in that it is required for – a necessary condition of – many, and perhaps all, of the other moral virtues. If the moral virtues are prior to the executive ones, this would entail that is a peculiarly foundational virtue – moral or executive. As Kundera notes, the most obvious candidates for those who have no power are animals. Some humans have no power, and the virtue of mercy will also underpin the virtue ethical case that can be mounted in support of them. But almost all animals are powerless relative to us. Certainly, the ones that we encounter in our everyday ‘civilized’ dealings are – the animals we eat, experiment on, and invite into our homes as companions are powerless relative to us. In his or her dealings with these powerless beings, the virtuous person will be guided by the virtue of mercy. And anyone who is not, is not, I think, a virtuous person.

6

Contractarianism and Animal Rights

1 Contractarianism and moral status

In this chapter, I shall argue that a strong – and perhaps the best – case for the moral claims of non-human animals can be made using the apparatus of contractarian or contractualist moral theory. The canonical version of contemporary contractarianism was supplied by John Rawls, in *A Theory of Justice* and subsequent writings. I am going to argue that contractarianism, of a form recognizably similar to that defended by Rawls, can be used to underwrite the moral claims of animals. In particular, it can be used to justify the claim that non-human animals possess moral *rights*.

For the sake of discussion, and for framing the argument I am going to develop in this chapter, I shall work with the account of rights employed by Tom Regan. Moral rights are valid claims all things considered. That is, moral rights are: (i) valid claims to a specific commodity, freedom, or treatment; (ii) made against assignable individuals who are capable of granting or withholding the commodity, freedom, or treatment; where (iii) a claim is valid if it is backed or entailed by a correct moral theory. The argument I am going to develop does not require this account of rights. But it is useful to have a fairly precise concept of rights at hand in order to develop this chapter's central claims. Unlike Regan, however, in this chapter the correctness of the moral theory in question will not rest on any obscure notion of inherent

value. On the contrary, the burden of this chapter is to show that a recognizable form of contractarian theory is a compelling candidate for the valid moral theory that underwrites the moral rights of non-human animals.

This claim will strike many as surprising. It is almost universally supposed that contractarianism is incompatible with animal rights. More precisely, it is almost universally assumed that contractarian approaches are unable to underwrite the granting of *direct* moral status to non-human animals, although they may be compatible with the granting of *indirect* moral status to the extent that non-humans bear certain relations to humans, the bearers of direct moral status. Put in the idiom of rights, it is both customary and important to distinguish between two sorts of rights, direct and indirect:

Direct rights. An individual I possesses a direct right R to a certain commodity, freedom, or treatment if and only if (a) I possesses R, and (b) the possession of R by I does not depend on the existence of rights possessed by any individual distinct from I.

Indirect rights. An individual I possesses an indirect right R to a certain commodity, freedom, or treatment if and only if (a) I possesses R, and (b) the possession of R by I does depend on the existence of rights possessed by an individual distinct from I.

There are various ways in which this distinction can manifest itself. For example, consider the claim that a dog possesses indirect rights. One way in which this might be so is that if I, as a possessor of direct moral rights, am emotionally attached to my dog so that harm to my dog would upset me in some way, then I might have a *prima facie* right to require that you do not harm my dog. My dog has no right not to be harmed by you, but I have a right that my dog not be harmed by you. And any harm visited upon my dog by you is an infringement not of my dog's direct rights (since he has none) but of mine. In this case, we can speak of the harm done

to my dog as an infringement of his indirect rights; rights that he possesses only in virtue of rights that I possess. In this case, then, the violation of my dog's indirect rights derives from the violation of my direct rights.

There is another well-known sense in which a dog might come to be a bearer of indirect rights. There is a view associated with Aquinas and Kant, among others, according to which a harm such as cruelty inflicted on my dog is wrong not because of the harm it does to my dog, but because of the deleterious effect it has upon the person who inflicts the harm.¹ Cruelty and callousness to non-humans is wrong not in itself, but because it tends to make the perpetrators cruel and callous and this can then go on to infect their dealings with other human beings. He who is hard in his dealings with animals becomes hard in his dealings with humans, or so the idea goes. Now, at this point, I am not at all concerned with whether this idea is correct (and, in particular, whether it gets the direction of causation right). The point is simply that this is another version of the indirect rights view. According to this view, my dog possesses indirect rights only in virtue of the existence of a distinct individual – a human – who possesses direct rights. It differs from the first case in that the violation of my dog's indirect rights here does not stem so directly from the violation of another's direct rights. Presumably, the person who is cruel to my dog does not thereby have his rights infringed upon. Nonetheless, what makes the cruelty to my dog wrong, on this view, is the tendency for it to lead to character traits which will extend to the person's interactions with humans, and therefore to violations of their direct rights. Ultimately, then, both ways of developing the indirect rights view make the possession of indirect rights by an individual dependent on the possession of direct rights by distinct individuals. Indirect rights can be possessed and violated only in virtue of the possession and violation of direct rights.

Put in terms of this distinction, the relation between contractarianism and animal rights is almost universally thought to look something like this. Firstly, contractarian theory can, when suitably developed, underwrite the possession of indirect moral status

to non-human animals. That it can do this is regarded as obvious. Secondly, contractarian theory cannot underwrite the possession by animals of direct moral status. Indeed, not only can it not underwrite this, it actually precludes this. Contractarianism is incompatible with the possession by animals of direct moral rights. This is also regarded as obvious.

The reason for this almost universal assumption is that non-human animals are (it is supposed) not rational agents, and contractarian approaches subsume, under the umbrella of moral consideration or concern, only rational agents. Thus, for example, according to Carruthers:

Morality is here [i.e., according to the contractarian approach] pictured as a system of rules to govern the interaction of rational agents within society. It, therefore, seems inevitable, on the face of it, that only rational agents will be assigned direct rights on this approach. Since it is rational agents who are to choose the system of rules, and choose self-interestedly, it is only rational agents who will have their position protected under the rules. There seems no reason why rights should be assigned to non-rational agents. Animals will, therefore, have no moral standing under Rawlsian contractualism, in so far as they do not count as rational agents.²

Carruthers endorses this conclusion and sees it, if anything, as a *strength* of contractarian approaches that they do not assign direct rights, or any other form of direct moral status, to non-humans. However, this view of contractarianism seems to be shared by both foes *and* friends of animal rights. Thus, Tom Regan, by far the most influential defender of the concept of animal rights, claims:

[I]t [Rawls's contractarianism] systematically denies that we have direct duties to those human beings who do not have a sense of justice – young children, for instance, and many mentally retarded humans.³

Regan shares with Carruthers the assumption that contractarianism, as represented by John Rawls, only applies to rational agents.

And since many, if not all, non-human animals cannot be regarded as rational agents in the relevant sense, contractarian approaches will fail to assign them direct moral rights.

I think this view can be questioned in several ways. Certainly, the all or nothing manner in which discussions of non-human rationality tend to take place is eminently questionable, on both theoretical and methodological grounds.⁴ This is not, however, the issue on which I am going to focus. Instead, I am going to argue that the assumption that contractarianism is incompatible with animal's possessing direct moral rights is, in general, indefensible – irrespective of whether the animal in question is rational or not. In particular, I shall argue that there is nothing in contractarianism *per se* which requires that the protection afforded by the contract be restricted to rational agents. The fact that the *framers* of the contract must be conceived of as rational agents does not entail that the *recipients* of the protection afforded by the contract must be rational agents. In fact, I shall argue that, for the most plausible versions of contractarianism, quite the opposite conclusion turns out to be true. In such versions, the recipients of the protection offered by the contract *must* include not only rational, but also non-rational, individuals.

This is a conclusion that most defenders of contractarianism will regard with incredulity. Even Rawls, whose theory the one in this chapter most closely approximates, thought of his account as not applicable to non-humans. I shall argue that this assumption, widespread and tenacious though it may be, is simply false. The reason so many people believe it to be true is because they fail to realize there are two very different versions of contractarian theory. One version rules out the possession of rights by animals, and non-rational agents in general. The other version, a far more plausible and influential version, can be used to underwrite the direct moral status of animals and other non-rational agents. The distinction between these two forms of contractarianism is, therefore, absolutely central to the arguments to follow; and so here it is we must begin.

2 Two forms of contractarianism

According to contractarian approaches to ethics, very roughly, the requirements of morality are determined by the agreements that humans make, or would make, to regulate their social interactions. There are, however, two importantly different types of contractarian theory, based on very different assumptions, and yielding very different moral principles. It is, however, rare to find these forms properly distinguished.⁵ The first form of contractarianism derives from Hobbes, and I shall accordingly refer to it as *Hobbesian* contractarianism. The second form derives from Kant, and will be referred to as *Kantian* contractarianism. The primary differences between Hobbesian and Kantian contractarianism concern their conceptions of (i) the *authority* of the contract: that is, the conditions that must be satisfied in order for the (hypothetical) contract to *bind* us or have authority over us, and (ii) in what this authority must be grounded.

According to Hobbesian contractarians, there is nothing objectively right or wrong either with the goals one chooses or the means by which one pursues these goals. In particular, there is nothing inherently wrong with harming others in order to achieve one's goals. However, while there is nothing inherently wrong with harming others, doing so may often be imprudent. Typically, I would be better off refraining from harming you if, in turn, you and every other person refrains from harming me. Thus, a convention that forbids deliberately harming people is mutually advantageous; we do not have to waste time, effort, money, and so on defending our own person and property, and it enables us to enter into stable, and mutually beneficial, co-operation. While deliberately harming another is not inherently or objectively wrong, it is nonetheless imprudent, and, therefore, wise to treat it as if it were wrong.

Thus, according to Hobbesian contractarianism, the basis of morality can be understood as a hypothetical contract consisting of mutually advantageous rules of conduct. The content of such conventions will be fixed by bargaining: each person will want the

resulting agreement to serve the dual purpose of protecting their own interests as much as possible while restricting their freedom as little as possible. While this bargaining never really took place, we can view this hypothetical bargaining over mutually advantageous conventions as the means by which a community establishes its *social contract*. The principles established by this imaginary bargaining process are to be obeyed not because it is inherently wrong to transgress them, but, ultimately, because it is irrational to do so. To this extent, and to this extent only, the hypothetical social contract can be thought of as yielding a moral code.

This conception of the function of morality yields, in turn, a conception of the *authority* of the contract. According to the Hobbesian form of contractarianism, the ultimate source of the contract's authority derives from the fact that we have implicitly agreed to it. All versions of contract theory, of course, accept that the contract is a hypothetical entity. So our implicit agreement in this context amounts to this: the contract embodies the rules that we either *have* endorsed or *would* endorse if they had been put to, and freely discussed by, us. The legitimacy, and therefore the authority, of the contract stems from our tacit agreement to the rules it embodies.

Our implicit assent to a hypothetical contract is not a brute fact, but is dependent on whether or not it is in our interest – our long-term, rational, interest – to endorse its rules. But, it will be in our long-term, rational, interest to endorse its rules only if the benefits we secure by doing so outweigh the restrictions on our freedom that such endorsement entails. The principal benefits we might secure are protection from those who might harm us and assistance from those who might help us. Therefore, there is no (Hobbesian) rationale for contracting with those individuals sufficiently weaker than you are to be in a position neither to help you nor to hinder you. We can refer to this as the *equality of power condition*.

Neither is it in your long-term, rational, interest to contract with those who are unable to understand the terms of the contract and therefore reciprocate in the ways required by it. Again, one would

be accepting restrictions on one's freedom without the possibility of getting anything of comparable importance in return. We can call this the *rationality condition*. Like the equality of power condition, the rationality condition is also entailed by the Hobbesian understanding of the contract and its authority.

Hobbesian contractarianism, therefore, exhibits the following characteristic structure: the *authority* of the contract is explained in terms of our *tacit agreement* to it; and our tacit agreement to the contract is explained in terms of our *rational self-interest*. But, in this context, rational self-interest makes sense only if those with whom we contract satisfy the *equality of power* and *rationality* conditions.

The second version of contractarianism is quite different. Hobbesian contractarianism uses the idea of a (hypothetical) social contract to ground morality, in the sense of providing a justification for a moral code and an explanation of why we should adopt the rules of conduct embodied in this code. The second version of contractarianism, however, uses the idea of a contract in a fundamentally different way. The contract idea, here, is used not as a method of grounding or justifying any particular moral code, but, rather, as a heuristic device in terms of which we can identify and express the principles embodied, often in a partially concealed or implicit manner, in the moral code that we have, for whatever reason, in fact adopted. For example, the contract device can be used in this way to express and reflect the idea of the equal moral status of persons, rather than as an account of how persons come to have moral standing. And the device can be used in this way to eliminate, rather than reflect, differences in the bargaining power of the contractors. This second version of the contract theory has its roots in the work of Kant, and we can therefore (with some reservations) refer to it as *Kantian contractarianism*.

Kantian contractarianism has a quite different conception of the authority of the contract. The Hobbesian contractarian sees the contract as *constitutive* of moral right and wrong: these are constituted or defined by the tacit agreements reached by rational

contractors of roughly equal power. The authority of the contract, therefore, derives from our tacit agreement to its conditions. Contained in the idea of Kantian contractarianism, on the other hand, is an at least minimal conception of moral *truth* or *objectivity* that is independent of the contract and the agreements reached by contractors. This is, in effect, the contractarian echo of Kant's notion of the *Moral Law*.

For the Kantian contractarian, the contract and contractual agreements will have authority to the extent that they embody, or at least approximate, moral truth or correctness. Our tacit acceptance of contractual arrangements does not constitute moral correctness. Rather, the arrangements themselves are subject to independent standards of correctness. The contract is not, as it is for the Hobbesian, a device that *constitutes* moral correctness or incorrectness, but, rather, one whose function is to help us *identify* or *reveal* what is morally right and wrong, and that has this status independently of the contract itself.

This approach is Kantian on several counts, ones that, I think, are collectively sufficient to justify the label. First of all, as I have pointed out, in the at least minimal, contract-independent, conception of moral truth it presupposes, the approach echoes Kant's idea of the Moral Law. Second, there is the idea that certain important (though not necessarily all) constraints on what is to count as right and wrong can be identified by way of an examination of the normative requirements for the attribution of moral predicates. The role of the contract is, fundamentally, to assist with the examination of these constraints. Third, unlike the Hobbesian alternative, the authority of the contract does not stem from the fact that it is in our interest to endorse its rules (which is why, it is assumed, we tacitly endorse them). This would be a *prudential* account of morality authority, and the Kantian form of contractarianism is, in an appropriate sense, *categorical*. Finally, there is, in a sense to be made clear, an important role for *intuition* in the development of the Kantian version of contractarianism, a role that has no real echo in the Hobbesian alternative.

For our purposes, however, the most important question is not the extent to which this alternative version of contractarianism can legitimately be labelled 'Kantian' but, rather, this: what is the source of the *authority* of the contract in this Kantian incarnation? That is: what makes the contract binding upon us? For Hobbes, of course, the authority of the contract lay in our tacit agreement to it. For the Kantian version, however, the authority of the contract is derivative upon the authority of the moral principles that it helps us uncover. And why are these binding? The answer is: if they are indeed morally correct principles, then obeying them is the right thing to do.

This different conception of the authority has, what is for our purposes, a crucial entailment: the *equality of power* and the *rationality* conditions play no essential role in the Kantian version of contractarianism. To see this, let us call the objective moral principles that the contract is to help us uncover the Moral Law. The contract does not determine the content of this Law, but is, rather, used as a heuristic device for allowing us to discover this content. Given that the function of the contract is, in this way, revelatory rather than constitutive, whether or not an individual who is deficient in point of power, or in point of rationality, or both gets included under the scope of morality is dependent only on what the Moral Law says: if it says the individual is in, he's in. If it says he's out, then he's out. So, whether or not an individual is to be included in the contract, and therefore within the scope of moral consideration, does not depend directly on his power or rationality but on whether the Moral Law says that he counts morally.

Depending on its content, the Law may, of course, specify that a person who is deficient in power or rationality does not count morally. Indeed, something like this may have been Kant's view.⁶ For Kant, a non-rational agent is not an end-in-itself, and therefore falls outside the scope of direct moral concern. However, the crucial point is that the exclusion of a non-rational individual is something for which additional argument is needed. It cannot simply be taken as given from the nature of the contractual situation. Here, the Moral

Law is the dog, and the contract is the dog's tail. The contract does not determine who does and who does not count morally.

The version of contractarian theory I am going to defend uses the contract not as a way of elucidating the content of any Moral Law in Kant's sense, but of the principle that, I have argued, provides the cornerstone of contemporary moral thinking: the principle of equal consideration. The role of the contract, as employed here, will be to help elucidate the content of this principle. And this principle, I shall argue, when properly elucidated and properly understood, will be seen to undermine both the equality of power and the rationality conditions. Therefore, I shall argue, the fact that the framers of the contract have to be regarded as rational and of roughly equal power does not entail that the recipients of the protection of the contract have to share these properties.

One may object that a form of contractarianism this 'Kantian' is not really contractarianism: that anything that is to count as contractarianism must, in Hobbesian manner, make the contract constitutive of moral right and wrong. If that is the case, however, much of what passes for contractarianism in recent decades is not contractarianism either. The version of contractarianism I am going to develop is essentially that of John Rawls, plus or minus a few twists here and there. However, most recent influential forms of contractarianism are unstable, and arguably untenable, mixtures of Hobbesian and Kantian forms of contractarianism.⁷ This is as true of Rawls's account as it is of others. In Rawls's version of contractarianism, I shall argue, we find a Kantian core surrounded by unexpurgated, unfortunate, and unnecessary elements of Hobbesianism. Much of the plausibility of Rawls's account stems from this Kantian core. And much of what is questionable about his account stems from the unnecessary Hobbesian residue. So one way of thinking about the goal of this chapter, then, is to exorcise Rawls of his Hobbesian demons, and so end up with a version of contractarianism that is truer to his underlying Kantian motivation and methodology than the one developed by Rawls himself. Once we have this, I shall argue, we shall see why it can be used to underwrite the (direct) moral claims of animals.

3 Contractarianism and animal rights: the orthodox view

The first task is to flesh out in a little more detail the reasons why contractarianism is thought to be incompatible with the possession, by non-humans, of direct rights. The underlying argument for this incompatibility seems to be of the following form:

- P1. According to contractarianism, moral rights and duties are dependent on the existence of an actual or hypothetical contract.
- P2. The framers of the contract and the moral rights and duties embodied therein have to be conceived of as rational agents.
- P3. Therefore, the contract and its embodied moral rights and duties apply only to rational agents.
- P4. Non-human animals are not rational agents.
- P5. Therefore, the contract and its embodied rights and duties do not apply to non-human animals.
- P6. Direct moral rights are possessed only by those individuals subsumed by the contract and its embodied rights and duties.
- C. Therefore, non-human animals do not possess direct moral rights.

This argument, I think, expresses the orthodox understanding of the relation between contractarianism and animal rights. The argument, of course, is compatible with non-humans being the bearers of indirect rights, but not with their possession of direct rights.

The argument is, of course, not deductively valid, and it would be unfair to present it as such and to criticize it for its failure in this regard. Nonetheless, there is still a large jump from P2 to P3. To claim that the framers of a contract must be conceived of as rational agents obviously does not entail that the recipients of the protection afforded by the contract must be similarly conceived. The argument can be rendered plausible, then, only if some justification for the move from P2 to P3 can be provided. However, all justifications

found in the literature seem to presuppose a Hobbesian form of the contract.

To take just one prominent example, David Gauthier focuses unashamedly on the usefulness that a contract would have for us.⁸ In order to regard a contract as in any way binding we have to recognize that it would be a good thing for us if we adhered to the contract. Adhering to the contract involves accepting certain restrictions upon one's freedom, and we will find this acceptable, and hence adhere to the contract, only if these restrictions allow us to obtain a good that outweighs the value of the freedom lost. Non-human animals, however, don't seem to fit very easily into this contractarian idea. They, being unable to understand the terms of the contract, cannot agree to abide by its principles. Therefore, it is argued, we would agree to accept restrictions on our freedom, but they do not. Therefore, we would lose something in agreeing to abide by the contract, and get nothing in return from them. Therefore, it is argued, non-human animals cannot be included as beings with whom we can meaningfully contract.

What is of interest at present is not the specific content of this justification, but the form it takes. The crucial assumption is that, if the contract idea is to work, then some account must be given of how the contract can be binding on us. Then, it is argued that the contract can be binding only if all the individual contractors agree to be bound in the same sort of way. Thus, any individual who cannot agree to be bound in the way specified by the contract cannot be meaningfully regarded as a contractor; and non-rational agents would provide a paradigm case of individuals who are not capable of being contractors in this sense. We can use this account to provide a supplemental premise:

- P2(a). Any individual who is not a contractor is subject neither to the conditions of the contract nor the protection afforded by the contract.

This, however, is a Hobbesian account of the authority of the contract. And this Hobbesian form of justification is not available to

the Kantian version of the contract. I shall argue that if we are consistently Kantian about the contract, then there is no way of bridging the gap between P2 and P3. The leap from P2 to P3 is, accordingly, an unjustified one. Therefore, Kantian contractarianism provides no reason for supposing that only individuals capable of framing the contract can be recipients of the protection offered by the contract. Indeed, when properly understood, I shall argue that the most plausible versions of Kantian contractarianism are committed to denying this claim.

4 Rawls and contractarianism

My target of an influential, if ultimately not quite consistent, form of Kantian contractarianism is supplied by John Rawls in *A Theory of Justice* and, more recently, *Political Liberalism*.⁹ The reason for focusing on Rawls is, of course, that he is (deservedly) the most influential of modern contractarians, and, consequently, any defence of contractarianism must effectively define itself in relation to Rawls's view. At the outset, however, one point of contrast should be noted. Rawls is primarily interested in political philosophy, and his application of contractarianism is used to determine the nature of what he calls the *basic structure of society*, that is 'the way in which the major social institutions distribute fundamental rights and duties and determine the division of advantages from social cooperation'.¹⁰ And by 'major social institutions' Rawls means the political constitution and the principal economic and social arrangements. I propose to use the contractarian idea in a somewhat broader sense as providing a general theory of morality; that is, as providing a framework for the assignment of moral rights and duties in general, and not just political rights of the sort discussed by Rawls. That is, the contractarian idea, as I propose to use it later in the chapter, will be conceived of as, in principle, being capable of providing us with general principles of morality, and not simply principles relating individuals to basic societal structures. While this differs in scope from

Rawls's view, this application of the contract idea is, of course, by no means idiosyncratic.

The ideas that form the conceptual heart of Rawls's contractarianism are those of the original position and the associated idea of the veil of ignorance. For Rawls, the way to think about what would be a just organization of society is to imagine what principles would be agreed to by people who were denied knowledge of certain facts about themselves. The people here find themselves in the original position, and the facts of which they have no knowledge are excluded by the veil of ignorance. The facts excluded by this veil can be divided into two sorts. Firstly, the occupants of the original position do not know their socio-economic position in society, nor do they know their own natural talents or endowments. Secondly, they do not know their own conceptions of the good; that is, given that there are alternative possible sets of beliefs about how one should live one's life, the occupants of the original position do not know which set of beliefs they will hold. The occupants of the original position are conceived of as rational. And while they do not suffer from envy, they are concerned to put themselves in as advantageous a position as possible after the lifting of the veil of ignorance.

Rawls claims that a person put in the original position would choose two principles of distributive justice:

First principle – Each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system for all.

Second principle – Social and economic inequalities are to be arranged so that they are both:

- (a) to the greatest benefit of the least advantaged, consistent with the just savings principle, and
- (b) attached to offices and positions open to all under conditions of fair equality of opportunity.¹¹

(He also defends various priority rules which are not directly relevant to our concerns).

It is important to realize, however, that the concept of the original position cannot, *by itself*, motivate these two principles. That is, Rawls, in fact, has two essential arguments for these principles of justice and not, as is commonly thought, one.¹² The first argument functions by contrasting his theory with what he takes to be the prevailing ideology concerning distributive justice – namely the ideal of equality of opportunity. The political system that embodies this ideal is referred to by Rawls as the system of *liberal equality*. Rawls argues that his theory (i.e., *democratic equality*) better fits our considered intuitions concerning justice, and that it more consistently spells out the very ideals of fairness that underwrite the prevailing ideology. I propose to call this the *intuitive equality argument*. The second argument defends the principles of justice by showing that they are the principles that would be adopted by rational agents in the original position. I shall refer to this as the *social contract argument*. Rawls has, of course, placed far more emphasis on the social contract argument, and this has led many people to overlook the intuitive equality argument. This, however, constitutes a serious oversight, since, as I shall try to show, the former is crucially dependent on the latter. Understanding the relation between the intuitive equality argument and the social contract argument is essential to understanding the way in which contractarianism can underwrite the attribution of rights to non-humans.

The intuitive equality argument

In broad outline, the basis of what I have called the *intuitive equality argument* looks like this:

- P1. If an individual I is not responsible for their possession of property P, then I is not morally entitled to P.
- P2. If I is not morally entitled to P, then I is not morally entitled to whatever benefits accrue from their possession of P.
- P3. For any individual I, there will be a certain set of properties $S = \{P_1, P_2, \dots, P_n\}$ such that I possesses S without being responsible for possessing S.

- C. Therefore, for any individual I, there is a set S of properties such that I is not morally entitled to the benefits which accrue from possession of S.¹³

In other words, if a property is *undeserved* in the sense that its possessor is not responsible for, or has done nothing to merit, its possession, then its possessor is not morally entitled to whatever benefits accrue from that possession. Possession of the property is a morally arbitrary matter and, therefore, cannot be used to determine the moral entitlements of its possessor. The argument also has a corresponding negative form, according to which, morally speaking, one should not be penalized for the possession of a property one has done nothing to deserve. It doesn't really matter which of the two forms we concentrate on. I shall focus on the positive form as described above.

Rawls believes that the above argument underlies the ideal of equality of opportunity which he identifies with the prevailing liberal orthodoxy. That is, the principle that *one is not morally entitled to benefits that accrue from properties one has done nothing to earn* is a principle which provides a conceptual foundation for the politics of liberal equality and its embodied ideal of equality of opportunity. Rawls endorses this principle. His case against the concept of equality of opportunity, as this is usually understood, stems not from a disagreement with the principle as such, but, rather, centres on the range of properties that should be regarded as morally arbitrary, and thus falling within the scope of the principle. It is a commonplace that being born into a certain position in society – in a particular social, racial, economic, or gender group – is an undeserved and, hence, morally arbitrary property. And, therefore, one should be neither benefited from nor penalized by possession of such a property. In other words, economic and social inequalities are undeserved, and, hence, it is unfair for one's fate to be made any better or worse by this sort of undeserved inequality. However, what the concept of equal opportunity, as this concept is understood in contemporary liberal cultures, overlooks

is that there are many more properties which are undeserved in the requisite sense. In particular, inequalities in natural talents or capacities are undeserved in precisely the same way as social, racial, economic, and gender properties. No one deserves to be born athletically gifted, stunningly handsome or with an IQ of 153, any more than they deserve to be born into a certain privileged class, sex, or race. Therefore, if it is unjust for someone to benefit from possession of undeserved social, racial, economic, or gender properties, then it must be equally unjust for them to benefit from possession of undeserved natural talents.

What is going on here is that we have a principle – the principle of equality of opportunity – which is embodied in contemporary liberal ideology, and is broadly accepted within this framework, but is not *consistently* implemented. Thus, Rawls's argument provides a more coherent and theoretically penetrating expression of the very assumptions which underlie the prevailing liberal view. Rawls's argument turns on the distinction – a distinction which will prove important in the arguments of later sections – between a principle being *embodied* in an ideology and that principle being *consistently adhered to* by proponents of that ideology. Rawls's point, in part, is that the former does not entail the latter. And where we have a dissonance between the embodiment of a principle and the consistent adherence to that principle, the moral philosopher's job, in part, is to point out, and hopefully rectify, this dissonance.

The social contract argument

Rawls's social contract argument runs as follows. We imagine a so-called *original position* whose occupants are behind a *veil of ignorance*:

No one knows his place in society, his class position or social status, nor does anyone know his fortune in the distribution of natural assets and abilities, his intelligence, strength, and the like. I shall even assume that the parties do not know their conceptions of the

good or their special psychological propensities. The principles of justice are chosen behind a veil of ignorance. This ensures that no one is advantaged or disadvantaged in the choice of principles by the outcome of natural chance or the contingency of social circumstances. Since all are similarly situated and no one is able to design principles to favour his particular condition, the principles of justice are the result of a fair agreement or bargain.¹⁴

The concept of the original position, and the associated idea of the veil of ignorance, will play a central role in this book's defence of the attribution of (direct) rights to non-human animals. Therefore, at this point it is essential to clear up one serious, and actually quite extraordinary, misunderstanding of these concepts that has become prevalent in recent years.

Many communitarian critics of Rawls have objected to the notion of an original position on the grounds that it entails a spurious metaphysical conception of the self. This claim is based on the idea that when the multifarious types of knowledge described above have been bracketed off, as demanded by the veil of ignorance, we are left with nothing but a radically *unencumbered* self. That is, we are left with a self which has its ends only contingently. Communitarians believe this is a false view of the self. It ignores the fact that the self is *embedded* or *situated* in existing social practices, and that these, in an important sense, define the self or constitute its identity as the particular self that it is. It makes no sense, then, on the communitarian view, to try and imagine a self in the original position. A self which occupied the original position would have had taken away from it precisely those features which constitute its identity; it would therefore have ceased to be a self. An unencumbered self, therefore, is radically unimaginable because the whole idea of a self occupying an original position is incoherent.¹⁵

This is, of course, not the place to enter into a discussion of communitarianism. Even without such discussion, however, it is not difficult to see that this sort of criticism is misguided. The concepts of the original position and veil of ignorance are neither

expressions of, nor do they entail, any metaphysical theory of the person. Rather, they constitute an intuitive test of fairness. Just as we might try to ensure a fair division of a pizza by making sure that the person who slices it does not know what piece he will get, so too we ensure a just distribution of rights by making sure those who are able to influence the selection process in their favour, due to their better position, are unable to do so. The cutter of the pizza does not know which piece he will get, therefore he cuts the pieces fairly; the distributor of rights does not know where in the distributive scheme he will fit, therefore he distributes justly. Similarly, the notion of the contract, in Rawls's hands, is not to be confused with any agreement – actual *or* hypothetical – but as a device for teasing out the implications of certain premises concerning people's moral equality. That is, the idea of the original position is used as a heuristic device to model the idea of the moral equality of individuals.

This being so, there is no dubious metaphysical conception of the self embodied in the concept of the original position. Firstly, the concept of the original position does not require that there could actually be a self, or selves, which inhabit the original position. That is, Rawls is not committed to the *metaphysical* possibility of occupants of the original position. Secondly, the concept of the original position does not even entail that it is possible to imagine the nature of occupants of the original position. That is, Rawls is not even committed to the *conceptual* possibility of a self existing behind the veil of ignorance. The reason he is committed to neither of these possibilities is because the original position and veil of ignorance are simply heuristic devices. Even in *A Theory of Justice*, Rawls is quite clear on the heuristic status of these concepts. He writes:

Some may object that the exclusion of nearly all particular information makes it difficult to grasp what is meant by the original position. Thus, it may be helpful to observe that one or more persons can at any time enter the original position, or perhaps, better, simulate the deliberations of this hypothetical situation, simply by reasoning in accordance with the appropriate restrictions... To

say that a certain conception of justice would be chosen in the original position is equivalent to saying that rational deliberation satisfying certain conditions and restrictions would reach a certain conclusion.¹⁶

And in later works, obviously mindful of the misunderstandings of his work on precisely this issue, Rawls is even more clear on the heuristic status of the concept of the original position. It is simply, as Rawls says, a 'device of representation', which serves as a means of 'public reflection and self-clarification'.¹⁷

One can 'enter' the original position, then, not by becoming a radically unencumbered self, but by reasoning in accordance with certain restrictions. More precisely, one can put oneself in the original position simply by imagining that one is without a certain attribute that one does in fact have, or without a certain conception of the good that one does in fact hold.¹⁸ This does not require that we imagine ourselves without *any* attribute or without *any* conception of the good. It simply requires that we be able to bracket these features of ourselves in a one-by-one piecemeal manner. Rawls indicates that he will go on speaking in terms of the original position because such talk is 'economical and suggestive', and brings out certain essential features one might otherwise overlook. Given the frequent and egregious misunderstandings occasioned by Rawls's use of this concept, one may legitimately wonder if this decision was wise. But be that as it may, the important point is that the concept of the original position, and the associated concept of the veil of ignorance, are both heuristic through and through.

In failing to recognize the heuristic status of the concept of the original position, one not only misunderstands Rawls's views, one also, I think, fails to grasp the power and originality of his thinking about justice. Correct understanding of the concept of the original position is so central to the case I shall make for animal rights that I shall return to the task of clarification in the next section. At present we must move on to consider the relation between the social contract argument and the intuitive equality argument.

The mutual dependence of the arguments

As was mentioned earlier, Rawls has placed much more emphasis on the social contract argument, and this has led many to overlook the intuitive equality argument. To do this, however, would be to fail to understand how Rawls's version of contractarian theory works. I shall try to show that the social contract argument cannot be understood independently of the intuitive equality argument (and nor, indeed, can the latter ultimately be understood in isolation from the former). The two arguments are, essentially, co-dependent and mutually reinforcing.

Rawls's defence of liberalism has been objected to on the grounds that he rigs the description of the veil of ignorance, and hence of the original position, in order to yield the principles of justice he requires (e.g., the difference principle).¹⁹ This sort of objection is, however, misconceived, since Rawls is perfectly willing to admit this. He recognizes that 'for each traditional conception of justice there is an interpretation of the initial situation in which its principles are the preferred solution'.²⁰ A description of the original position is, in part, a specification of which properties are to be excluded behind the veil of ignorance. There are many possible descriptions of the original position that are compatible with the goal of creating a fair decision procedure, and the difference principle would not be chosen in all of them. So, in order to determine which principles would be chosen in the original position, we first need to know which description of that position to accept. And, according to Rawls, one of the grounds on which we choose a description of the original position is that it yields principles we find intuitively acceptable. That is, one important way of justifying a description of the original position is that it yields the sort of principles which emerge from the intuitive equality argument. This is so because it is precisely this argument which is based on the principles embodied in our contemporary liberal ideology.

In deciding on the preferred description of the original position, we 'work from both ends'. This means that if the principles that are

yielded by a given description of the original position do not match our convictions of justice, as expressed in the intuitive equality argument, then we have a choice:

We can either modify the account of the initial situation or we can revise our existing judgments, for even the judgments we take provisionally as fixed points are liable to revision. By going back and forth, sometimes altering the conditions of the contractual circumstances, at others withdrawing our judgments and conforming them to principle, I assume that eventually we shall find a description of the initial situation that both expresses reasonable conditions and yields principles which match our considered judgments duly pruned and adjusted. This state of affairs I refer to as reflective equilibrium.²¹

The latter state of affairs is described as an equilibrium because the principles yielded by the original position and the judgements yielded by the intuitive equality argument ('duly pruned and adjusted') coincide; and it is reflective since we now know to what principles our intuitive judgements of equality conform.

It is important to realize that Rawls, in this passage, is advocating working from *both* ends. Not only can our description of the original position be modified by our intuitive judgements of equality, but so too can our intuitive judgements of equality be modified by our description of the original position. The relation, in other words, is genuinely dialectical. Failure to appreciate this point can often lead to an ultra-conservative interpretation of Rawls, according to which our description of the original position is wholly at the mercy of our intuitive judgements of equality, themselves seen as not subject to this kind of review or modification. This interpretation of Rawls, I think, robs his position of much of its power and distinctiveness. And, in any event, it is far from Rawls's notion of reflective equilibrium. Indeed, it seems much more akin to what we might call *unreflective* equilibrium. This sort of unreflective equilibrium, I shall argue, lies at the heart of the view that contractarianism does not provide an adequate foundation for animal rights.²²

And, once again, in deciding when our intuitive but unreflective judgements concerning justice should be overridden, the intuitive equality argument is crucial.

To see this, recall how Rawls was able to override the intuitive but unreflective judgements underlying the politics of liberal equality, identified by Rawls as the prevailing liberal ideology. The prevailing justification for economic distribution in our society is based on the idea of 'equality of opportunity'. Inequalities of income and prestige and so on are assumed to be justified if and only if there was fair competition in the awarding of the offices and positions that yield those benefits. This conflicts with Rawls's theory, for while Rawls also requires equality of opportunity in the allocation of positions, he denies that the people who fill the positions are thereby entitled to a greater share of society's resources. A Rawlsian society may pay such people more than average, but only if it benefits all members of society to do so. Under the difference principle, people only have a claim to a greater share of resources if they can show that it benefits those who have lesser shares. Thus, Rawls's theory conflicts with what passes for common sense in capitalist societies. What motivates this common-sense view is the idea that it is fair for individuals to have unequal shares of social goods if those inequalities are earned and deserved by the individual and, conversely, that it is unfair for individuals to be disadvantaged or privileged by arbitrary and undeserved differences in their social circumstances. As Rawls points out, however, there is another source of undeserved inequality that this argument ignores. While it is true that social inequalities are undeserved, it is also true that inequalities in natural talents are equally undeserved. No one deserves to be born handicapped, or with an IQ of 70, any more than they deserve to be born into a certain underprivileged class, race, or sex. Therefore, distributive shares should not be influenced by these factors either. What is going on here is that Rawls is using the intuitive equality argument to undermine a widely accepted, indeed common-sense, conception of just distribution. This common-sense idea of just distribution is no doubt intuitive – after all,

many people intuit it – but not sufficiently reflective. And to rely on such a principle and use it to determine the interpretation of the original position would not be a case of reflective equilibrium, it would, as we might say, be a case of unreflective equilibrium. Thus, what determines whether an intuition of justice is a reflective intuition or not is the consistent application of the intuitive equality argument. This argument, therefore, plays a central role in determining the correct description of the original position, and, therefore, the principles of justice which are derived from this.

Later in the chapter, I shall argue that many of the arguments against extending a Rawlsian conception of justice to non-humans, remarks issuing from Rawls as well as others, are based on unreflective intuitions; intuitions not compatible with the consistent application of the intuitive equality argument. Since it is the consistent application of this argument that *should* determine which description of the original position we employ, these unreflective intuitions can play no role in determining this description. Now, however, it is time to see how a Rawlsian version of contractarianism can be used to underwrite the attribution of rights to non-humans. The first essential stage is to return to the task of clarifying the concept of the original position.

5 The original position revisited

In order to understand how the concept of the original position can provide a logical foundation for attribution of rights to non-humans, it is essential to remove certain serious misunderstandings that surround this concept. This will be the task of this section.

The task of clarification began in the previous section when replying to communitarian criticisms of Rawls. I argued that Rawls was not committed to a view of the self as essentially unencumbered. That is, Rawls was committed neither to the metaphysical nor even to the conceptual possibility of an unencumbered self. This is actually part of a wider issue.

The crucial point is this. The concept of the original position, and the associated concept of the veil of ignorance are simply heuristic concepts, or, as Rawls puts it, 'devices of representation'. The original position should *not* be thought of as any kind of state of affairs. That is, the concept of the original position is not a descriptive concept in the sense that it does not function to describe a situation or state of affairs. And this is true whether the envisaged state of affairs is conceived of as actual *or* as merely logically possible.

It is fairly clear, of course, that the concept of the original position does not function to describe any *actual* situation. Viewed in this way, the function of the concept would be to make an extremely implausible empirical claim; and no one, it seems, would want to suppose that this is indeed its function. However, even though the function of the concept is seen not to be descriptive of an actual state of affairs, many have thought that it does function to describe another type of situation. That is, many have thought that the function of the concept is to pick out a hypothetical, imaginable, or logically possible situation or state of affairs. I want to argue, on the other hand, that the concept does not function to pick out *any* state of affairs, whether actual or logically possible. Therefore, the concept of the original position does not entail that it is possible to *imagine* a radically unencumbered self of the sort that could occupy a hypothetical original position. Nor does it entail the *logical possibility* of a radically unencumbered self occupying an original position. That is, the concept of an original position entails neither the imaginability or logical possibility of an unencumbered self nor the imaginability or logical possibility of a position in which such a self could meaningfully be thought to be.

What the concept of the original position describes is not a possible state of affairs nor an imaginable one, but, rather, a certain type of reasoning process. This process of reasoning looks something like this: 'As a matter of fact, I have property P. But what if I did not have P? What principles of morality would I want adopted if I didn't have P?' One 'enters' the original position, in the only meaningful sense in which one can be said to enter it, when one engages

in a reasoning process of this type. And, Rawls's talk of a self occupying an original position is simply a way of adverting to a person who is restricting his or her reasoning about morality in accordance with the above sort of schema. That is, being in the original position is not a matter of being in a logically, metaphysically, or physically possible situation. It is simply a matter of allowing one's reasoning about morality to be guided by the above sorts of restrictions. Two important clarifications are in order here.

Firstly, given this understanding of the concept of the original position, there is no requirement that to be in this position one must have bracketed *all* one's properties. That is, in order to occupy the original position, one does *not* need to ask oneself the following sort of question: 'What moral principles would I want adopted if I had none of the properties I now, in fact, know myself to have?' Imagining oneself without any properties would, of course, be tantamount to imagining an unencumbered self. However, this is not required. All that is required is that one be able to bracket, or suspend belief in one's possession of, each individual property in a piecemeal, one-by-one, manner. The process is akin to repairing the Ship of Theseus while still at sea.²³ In order to avoid being partial to a particular distribution of moral principles on the basis of one's possessing a given property, one simply has to imagine not having that property and asking oneself what moral principles one would like to see adopted in that situation. Identification of the most adequate set of moral principles, then, is simply a matter of collating the results from these sorts of piecemeal inquiries.

Secondly, we need to observe an important distinction between what we can call *imagining that* and *imagining what it would be like*. Suppose our moral reasoner, for example, had the property of being male. In order to 'enter' the original position, he would have to reason in the following sort of way: 'Suppose I didn't have the property of being male. What principles of morality would I like to see adopted in that situation?' Since the person, *ex hypothesi*, has the property of being male, this is a case where he imagines *that*

he is not male. However, there is nothing in this procedure which requires him to imagine *what it would be like* to not have the property of being male. Imagining that you don't have a particular property and imagining what it would be like not to have that property are two very different things.

This distinction, of course, derives from a distinction between two different types of knowledge. On the one hand there is factual knowledge, often referred to as knowledge by *description*, that is, knowledge that a particular description can be applied to (or withheld from) a given object. On the other hand, there is knowledge by *acquaintance*, knowledge which is constituted by direct personal awareness or consciousness of something. The distinction is, of course, familiar from the work of Bertrand Russell.²⁴

The crucial category of knowledge/imagination for the purposes of the original position is knowledge or imagination *that*. One needs to be able to imagine that one does not have a particular property one in fact does have; one does not need to be able to imagine what it would be like to not have that property. Firstly, the latter demand would, in many cases, be extremely difficult, if not impossible, to satisfy; I might have no idea what it was like to be female. Secondly, even if the condition was possible to satisfy, there would be severe difficulties in actually determining when one had in fact satisfied it. I, for example, would have no way of knowing if my imaginative exercises had in fact succeeded in yielding to me the awareness of what it is like to be female.

Fortunately, being 'in' the original position does not have to involve the ability to imagine what it would be like to lack a given property. The notion of imagination employed in 'entering' the original position is an utterly minimal one: it amounts to *supposing* that one does not have a property that one, in fact, has. In the case of my lacking the property of being male, for example, I do not have to be capable of imagining what it would be like to be female in order to deduce what moral principles I would like adopted in that situation. All that is required is that I know certain pertinent *facts* about women. The relevant facts, here, would include things

like preferences and how a hypothetical moral or political arrangement would impact on those preferences. Thus, the fact that a given preference P might be so far removed from the preferences I in fact possess that I find it difficult or impossible to imagine having P does not undermine the validity of the original position. All that is required for me to be in the original position is that I know *that* a person has a given preference, not that I know what it is like to have that preference. This point has fairly obvious implications for the possibility of bringing non-human animals under the scope of the protection afforded by the contract. The practice of attributing preferences and other types of mental states to non-humans will be discussed at length in the next chapter.

In addition to these clarifications, there is also one crucial corollary of the above understanding of the concept of the original position. The corollary is this: *the contract does not necessarily involve distinct individuals contracting with each other*. The idea of different agents contracting with each other is not an essential part of the idea of the original position. The original position is perfectly compatible with a construal whereby we imagine various agents contracting with each other behind a veil of ignorance. However, neither the multiplicity of agents nor the notion of contracting is essential to the idea of the original position.

To see this, consider the following scenario. Imagine that *metempsychosis*, transmigration of the soul, is in fact true. And suppose, at some time when you are in between souls, God says to you: 'I am not going to tell you who or what you are going to be in your next life. However, I shall allow you to choose what moral principles you would like to see adopted in whatever world it is you are going to inhabit.' This way of setting up the original position commits you neither to the existence of God, nor to a dualist view of the person, nor to the transmigration of souls. It is simply a way of letting your reasoning and resulting moral choices be guided by certain restrictions. That is, this is simply another way of setting up the original position, and putting in place restrictions on one's reasoning about oneself that effectively constitute this position. The veil of

ignorance is, in this case, implemented by the fact that you do not occupy a body and God will not tell you which body you are going to occupy next. Thus, whatever restrictions on your knowledge are thought to be involved in Rawls's version of the original position can also be mirrored in this metempsychotic version. The crucial point, of course, is that any possibility of partiality is removed by your ignorance of your position in society; your conception of the good, and so on.

When we view the original position in this way, it is fairly clear that the original position can be occupied by one person alone. There is no need to view the position as one in which a multiplicity of rational agents contract or hammer out agreements amongst themselves. While it is perfectly consistent to imagine the original position as one in which a collection of distinct individuals contract among themselves in this way, this is not essential to the setting up of the position. One person denied any knowledge of him- or herself satisfies the conditions of the original position in an equally legitimate way. That is, what is crucial to the original position is the idea that an individual is denied all particular knowledge about him- or herself (or whatever subset of particular knowledge is deemed relevant by the intuitive equality argument), and is forced to choose principles of morality on this basis. The idea of distinct individuals denied such knowledge contracting with each other to choose these principles is an additional, and non-essential, element.

6 The original position and animal rights

I have argued that the Hobbesian version of contractarianism is defined by the following structure: the *authority* of the contract is explained in terms of our *tacit agreement* to it; and our tacit agreement to the contract is explained in terms of our *rational self-interest*. Rational self-interest, in turn, presupposes that those with whom we contract satisfy the *equality of power* and *rationality* conditions. Now, however, we are in a position to see that Rawls's

version of the contract, when properly understood, can accept neither the equality of power nor the rationality conditions.

Consider, first, the equality of power condition. According to this, contractual bargaining can take place only between contractors of roughly equal power. No version of contractarianism that is based on the original position can accept the equality of power condition – not as a *requirement* on the possibility of the contract. The original position does not even require the multiplicity of contractors – it does not require that the contractual situation contain more than one contractor. Therefore, it can hardly require that the contractors be of roughly equal power. (We could, of course, claim that a solitary contractor is of equal power with him- or herself. But it is difficult to imagine a move more reeking of desperation.) The equality of power condition is an unexpurgated version of Hobbesian contractarianism that has no place in its Kantian counterpart (when this is properly understood).

Consider, now, the rationality condition. When the connection between the description of the original position, the principles derivable from the position, and the intuitive equality argument is properly grasped, there is no reason to think that the bearers of the rights derivable from the original position are restricted to rational agents. The fact that it is (ideally) rational agents who, in the original position, are responsible for formulating the principles of morality does not entail that these principles subsume, or apply to, only rational agents. Indeed, given the nature of the intuitive equality argument, and the dependence of the description of the original position on this argument, it is clear why this claim should be rejected.

To see this, recall, firstly, the intuitive equality argument. The argument, in essence, runs as follows: If a property P is undeserved, in the sense that one is not responsible for possessing it, then it is morally arbitrary and one is not morally entitled to whatever benefits stem from the possession of P. However, rationality seems to be an undeserved property if any property is. A person plays no role in deciding whether or not she is going to be rational; she either is or

she is not. The decision is not hers, but nature's. Therefore, according to the terms of the intuitive equality argument, it is a morally arbitrary property, and one is not morally entitled to its possession. Therefore, also, one is not morally entitled to whatever benefits accrue from its possession. Therefore, to restrict the beneficiaries of the protection afforded by the contract to rational agents would be to contravene the intuitive equality argument. But it is the results of this argument which, in large part, determine the description of the original position. That is, it is the intuitive equality argument that provides the justification for excluding properties behind the veil of ignorance. Why do we exclude race, sex, natural intelligence, and so on behind the veil of ignorance? We do it precisely because these properties have been shown by the intuitive equality argument to be morally arbitrary. However, the principles of morality we derive from the original position will, of course, be a function of the properties we exclude behind the veil of ignorance. Therefore, given rationality is a property that is undeserved, it must, by the intuitive equality argument, be excluded behind the veil of ignorance. Therefore, the restriction of the beneficiaries of the contract to rational agents is one that we cannot legitimately apply. In the original position, one of the things we cannot know is whether we are going to turn out to be rational agents. And our principles of morality have to be chosen accordingly.

It is true that we sometimes speak of a person cultivating their rationality, or of endeavouring to do the rational thing in a given situation. And this may lead one to think that possession of rationality is something over which we have control, or even have to earn. However, this is not the sense of rationality that is relevant to the notion of moral consideration. This point is made quite forcefully by Rawls himself. No one, presumably, would want to claim that the more rational a person is, the more rights they have. Rationality, in the only sense possibly relevant to determination of moral rights, is what Rawls calls a *range property*.²⁵ For example, the property of being on the interior of the unit circle is a range property of points on a plane. All points inside the circle have this property although

their coordinates vary within a certain range. And they all have this property equally. It is rationality conceived of as this sort of range property that is employed by Rawls. We all possess rationality, and we all do so equally, even if some of us do better on IQ tests than others. And conceived of in this way, it is clear that our possession of rationality is not something over which we have any control. Our possession of this property depends on nature, and not on our own decisions and actions. It is, therefore, a morally arbitrary property in Rawls's sense.

The claim that rationality should be excluded behind the veil of ignorance is, I think, one that corresponds to common sense. None of us can be certain how much of our lives will be lived as a rational agent. We all know that during the early years of our lives we were not rational agents. We suspect that the later years of our lives might return us to the same state. And at every moment in between, the possibility of irreparable brain trauma due to accident or illness dogs our tracks. What would be positively irrational is to refuse to acknowledge these possibilities or eventualities, and so make no provision in the original position for them. That is, it would be irrational to choose a moral system that excluded the rational, given the near certainty that we will one day number among them, and the possibility that this might happen at any time.

Given the interdependence of the intuitive equality argument and the social contract argument, it seems that knowledge that one is a rational agent should be bracketed in the original position. This is what the intuitive equality argument tells us. This also coheres with common sense. It is also worth noting that the claim that knowledge of one's own rationality should, in the original position, be bracketed coheres much better with one of Rawls's ways of characterizing the original position as one in which the participants have knowledge of all general principles of psychology, sociology, economics, and the like, but *no* particular knowledge about themselves.²⁶ Since knowledge that one will be a rational agent is an obvious case of particular knowledge of the properties of oneself, it seems that this must be bracketed in the original

position. And if one does not know that one will be a rational agent, then, if Rawls is correct, one will, in the original position, inevitably formulate principles that take this into account. And at the very least, this would bring non-rational beings under the scope of the difference principle.

Therefore, when the relation between the social contract argument and the intuitive equality argument is correctly understood, it is seen that knowledge of one's own rationality must, for the sake of consistency, be bracketed in the original position. Hence, there is, or at least should be, nothing in Rawls's position which entails that non-rational creatures fall outside the sphere of justice. Similarly, there is nothing in the concept of the original position which entails that non-human animals fall outside the sphere of morality. On the contrary, once it is understood that what moral principles we can deduce from the original position depends on the description we give of that position, and once we understand that what we regard as an adequate description of this position derives from the consistent application of the intuitive equality argument, then we must allow that the principles of morality apply equally to both rational and non-rational individuals.

In fact, once the connection between the intuitive equality argument and the principles derivable from the original position is made clear, it seems that knowledge that one is a human being must also be bracketed in the original position. The property of being human is, again, something over which we have no choice. The property is undeserved in the sense that we are not responsible for possessing it. Therefore, according to the intuitive equality argument, the property is as morally arbitrary as the property of belonging to a given class, race, or gender. It is something over which we have no control. Therefore, according to the intuitive equality argument, we are not morally entitled to whatever benefits accrue from possession of this property. Therefore, given that the considerations underlying the intuitive equality argument are partly constitutive of the description we give of the original position, knowledge of one's human status is knowledge that should be bracketed in the

original position. Therefore, if the above arguments are correct, the sphere of morality should not be restricted to human beings.

7 Hobbesian remnants

According to Hobbesian versions, the contract is constitutive of moral right and wrong. In the hands of Rawls, the contractual apparatus serves a quite different function. Rawls's theory is based on a *pre-contractual commitment* to the idea of justice as fairness or impartiality. The contractual apparatus is used to elucidate, and render consistent ('prune and adjust'), the content of our intuitive concept of impartiality. In this, Rawls's approach is thoroughly Kantian in the sense explained above.

The Kantian component of Rawls's account provides us with a strikingly simple case for the moral claims of animals, where these are understood as claims of justice. The intuitive equality argument justifies the exclusion behind the veil of ignorance of underserved properties – properties that the subject has done nothing to earn or merit. These, according to the intuitive equality argument, are morally arbitrary properties, and so cannot determine the moral entitlements of a subject. However, it seems both the level of *rationality* and of species *membership* are undeserved properties in the relevant sense. Therefore, they should be excluded behind the veil of ignorance. In the original position, one should not know whether one is a rational agent, and one should not know the species to which one belongs.

Rawls is widely thought to reject the idea that animals can be incorporated into a theory of justice. I am going to attack this widespread assumption. My attack is three-pronged. The *first* prong tries to show that Rawls's alleged rejection of the idea that animals can be incorporated into a contractarian-based theory of justice is, in fact, far more equivocal than is generally acknowledged. This prong is based on an examination of what Rawls actually does say (in *A Theory of Justice*) about animals. The *second* prong of the

attack is based on drawing a firm distinction between the claims Rawls *in fact* makes about the status of animals and what claims are actually *required* or *permitted* by his contractarian model of justice. The aim of this aspect of the attack is to show that even if Rawls is as dismissive of the claims of animals as many people believe he is, this dismissal is in no way mandated by his contractarian theory. Indeed, his theory, properly understood, shows just why animals should be incorporated into a contractarian theory of justice. The *third* prong of the attack is a diagnosis of Rawls's failure to understand this entailment of his (own) theory. This failure, I argue, stems from the presence in his account of unexpurgated and unnecessary elements of Hobbesianism; elements of which Rawls could never entirely free himself.

First of all, let us examine the textual evidence for Rawls's supposed rejection of the idea that animals can be included under the scope of a theory of justice. Rawls claims that it is *moral persons* that are entitled to justice, where:

Moral persons are distinguished by two features: first they are capable of having (and are assumed to have) a conception of their good, as expressed by a rational plan of life; and second they are capable of having (and are assumed to acquire) a sense of justice, a normally effective desire to apply and to act upon the principles of justice, at least to a certain minimum degree.²⁷

However, even if we assume that animals are not moral persons in this sense, it follows that they are not owed justice in Rawls's sense only if being a moral person is a *necessary* condition of being owed justice. However, Rawls is quite clear that being a moral person is only a *sufficient* condition of inclusion under the scope of a theory of justice:

We see, then, that the capacity for moral personality is a *sufficient* condition for being entitled to equal justice. Nothing beyond the essential minimum is required. *Whether moral personality is a necessary condition I shall leave aside.*²⁸

Therefore, for Rawls, the failure of animals to count as moral persons does not, contrary to popular belief, automatically disqualify them from falling under the scope of a theory of justice. At several points, however, Rawls does seem to explicitly claim that animals are not entitled to equal justice:

Our conduct towards animals is not regulated by these principles, or so it is generally believed.²⁹

Presumably this excludes animals; they have some protection certainly, but their status is not that of human beings.³⁰

While I have not maintained that the capacity for a sense of justice is necessary in order to be owed the duties of justice, it does seem that we are not required to give strict justice anyway to creatures lacking this capacity.³¹

The hesitation in these claims is, however, obvious in the qualifications he uses to make them: 'or so it is generally believed', 'presumably', 'it does seem' – hardly ringing endorsements! The claims seem to have the status of (what Rawls himself would call) *unreflective intuitions* – intuitions embodied in common sense, but not duly 'pruned and adjusted' by the sort of deliberations required to produce a mature conception of justice, or the sort associated with a genuine attempt to reach reflective equilibrium.

More importantly – and this is the second prong of the attack – these unreflective intuitions are incompatible with Rawls's theory. It is the intuitive equality argument that determines which properties are to be excluded, at least provisionally, behind the veil of ignorance. And, by that argument, rationality and species membership – and, for that matter, moral personality – should be excluded on the grounds that they are unmerited. The intuitive equality argument, on my reading of it, provides only a provisional basis for exclusion of attributes behind the veil of ignorance. Nevertheless, it does provide an initial basis for exclusion, and any subsequent decision to rescind this exclusion requires

arguments – arguments that show why the decision to overturn the deliverances of the intuitive equality argument should be made. Rawls never provided such arguments. Citing moral personality would, of course, be question-begging, since this is precisely one of the attributes that the intuitive equality arguments entails we should bracket.

The question, then, is why was Rawls unable to see this entailment of his own theory? Here we arrive at the third prong of the attack. Rawls's suggestions that animals be excluded from the scope of a theory of justice are expressions of a residual Hobbesianism: they reflect unexpurgated, unnecessary, and unwelcome elements of a Hobbesian outlook of which Rawls could never quite free himself.

One Hobbesian principle that Rawls never rejected was the idea that the contractual situation involves at least rough equality of power between contractors: the equality of power condition. A rough equality of power between contractors is part of what Rawls called the *circumstances of justice*. However, as we have seen, this condition is no part of a Kantian form of contractarianism, properly understood. One way of thinking about the original position is as the coming together of a group of unencumbered individuals who then thrash out the terms of their association under the requisite conditions of ignorance. However, as we have seen, the original position is merely a heuristic device, and compatible with many such pictures. The metempsychotic picture, described earlier, is one in which the Hobbesian underpinnings have been more adequately expunged. You are a disembodied soul, between bodies. God informs you that while He will not allow you to choose which body you are going to occupy, He will allow you to choose the nature of the society in which you are going to live. That is, He will allow you to choose the principles of justice embodied in, and adhered to by members of, that society.

To adopt this *metempsychotic* version of the contractual situation, one needs to believe neither in God nor the possibility of disembodied souls – any more than one need believe in

unencumbered selves to adopt the Rawlsian version. To put oneself in the original position is just to allow one's reasoning to be governed by certain restrictions – restrictions on the knowledge one has of oneself and one's place in society. In this *metempsychotic* version of the contractual situation, one finds oneself behind a veil of ignorance of a familiar Rawlsian sort. Since you don't know who you are going to be, in choosing what is best for yourself you also choose what is best for everyone.

In this version of the contractual situation, there are no other contractors. It might be objected that one can know what is best for everyone only by thrashing this out with other contractors. But this was, in fact, always a red-herring. Any individual in the Rawlsian contractual situation is assumed to be ideally rational and also in possession of all pertinent general knowledge about human beings, their needs, desires, and so on. Such an individual would make an ideally rational choice for everyone. The presence of other contractors was, always, just a contingent feature of the contractual situation. It is worth noting that this is, in effect, acknowledged by Rawls. In a passage cited earlier, he writes:

Thus, it may be helpful to observe that *one or more persons* can at any time enter the original position, or perhaps better, simulate the deliberations of this hypothetical situation simply by reasoning in accordance with the appropriate restrictions ...³²

One or more persons can enter the original position by reasoning in accordance with the appropriate restrictions – that is, requisite ignorance concerning one's abilities, situation, and so on. The presence of more than one contractor is, even for the Rawls of *A Theory of Justice*, a contingent feature of the contractual situation.

However, if the original position need not contain more than one contractor, then the equality of power condition makes little sense in this Kantian context – unless one wants to resort to the desperate measure of talking of equality of power with oneself. If a contractual situation can obtain with only one person in it – if

the presence of more than this one person is a contingent feature of the situation – then there is little content to the idea of equality of power between contractors. So why does Rawls adhere to it? My diagnosis is that it is an incongruous remnant of the Hobbesian vision of the contract as a means of securing personal advantage. It is there because we tacitly assume the contract makes no sense – and cannot be binding – unless we get something out of it; at least as much as we put in, in fact. This is a Hobbesian idea, not a Kantian one.

Consider, now, the rationality condition. According to this, only a creature capable of framing a contract can be a recipient of the protection offered by that contract. Underlying the claim seems to be the following line of thought. First, you must be assumed to be rational in the contractual situation – that is, in the original position. This is obviously true: only a rational agent is capable of choosing a system of moral rules and requirements. Second, therefore, since you know you are rational in the original position, it would make no rational sense to make provision for something that is not. However, this is an extraordinary *non sequitur*. And even if we restricted the scope of morality to humans, as Rawls tries to do, it would *still* be an extraordinary *non sequitur*. If we assume – as we have no business doing – that in the original position you know you are human, then, knowing all general laws and facts about the world, you would know that you are a fragile creature who might, at any time, become a non-rational agent. (You might be hit by a bus just after you have concluded your contractual deliberations). It would be irrational to not make at least some provisions for this possibility. And it would be even more irrational to make no provisions for the strong likelihood that you will, through the ravages of time, become, once again, a non-rational.

The idea that only creatures capable of framing a contract can be recipients of the protection afforded by the contract is one that has no place in a Kantian version of contractarianism. What is doing the work bridging the gap between framers and recipients of the contract is a Hobbesian idea that is predicated on, and only

makes sense in the context of, a conception of the *authority* of the contract as deriving from our tacit acceptance of it, and our tacit acceptance of it as grounded in prudential reasons. That is, what is doing the work is the idea that it makes sense only to contract with those who are capable of helping you or hindering you. Failing this, there is no reason to accept, even tacitly, the contract. And without tacit acceptance the contract has no authority – it does not bind us. These are all Hobbesian ideas, not Kantian ones.

8 Other objections

In the literature there are two further common objections to using contractarianism to underwrite the notion of animal rights.

The first of these runs as follows. Once we start extending the scope of morality to include non-humans, there is, in principle, no limit to this extension. If we are willing to accord moral rights to non-humans, why not accord it to plants, even to inanimate objects? After all, in addition to rationality and species membership, sentience is also an undeserved property: whether or not we are sentient creatures is something over which we have no control. Therefore, on the above argument, sentience should also be excluded behind the veil of ignorance. The worry is that this exclusion, in one way or another, makes no sense.

One way in which it might make no sense is that it yields the supposition that I might turn out to be a rock or a tree. Of course, this supposition does make no sense. But neither, for that matter, does it make sense to suppose that I – the very same person – might have turned out to be a dog or cat. Nor does it make sense to suppose that I – the same person – might have turned out to be a woman rather than a man. Indeed, neither does it make sense to suppose that I – the same person – might have turned out to be a different man (at least if we assume that this other man has different parents).³³ It is still the case that every possible criterion of personal

identity will be violated by this supposition. If you were another human being you would have a different body and brain, so body- and brain-based criteria of personal identity would not apply. Also, since you would have a different set of memories and other psychological states, accounts of personal identity based on psychological continuity would also be inapplicable. Also, unless you turn out to be a sibling – and presumably we don't want to restrict the contractual situation to them – you will have different parents. And this, as Kripke has shown, makes it metaphysically impossible for you to be the same person.

It might be thought that this renders Rawls's position incoherent. It does not, however, for the simple reason that, as emphasized in previous sections, the original position is not a place where one can be, it is simply a certain type of reasoning process. That is, the function of the original position is not to describe a logically possible situation. It is to facilitate the shaping of one's moral reasoning in accordance with certain restrictions constitutive of the moral point of view. Therefore, talk of persons being 'in' the original position is metaphorical through and through. And, therefore, questions of the identity of persons in the original position with those outside it simply do not arise. To suppose that that they do is to misunderstand the nature and status of the original position.

Another worry turns on the idea that if we put so many properties behind the veil of ignorance, then it is unclear what the purpose of morality is any more. If I have to take into consideration the 'possibility' that I might not even be sentient, then too many things would be included within the scope of moral consideration. I would be choosing a world for trees and rocks, as well as cats and dogs, men and women. However, this worry is also misguided. The 'possibility' – and note that the word is placed in scare quotes to indicate that it is not a genuine possibility at all – that I am a rock or tree is one that would have no impact on my moral decision-making: for the simple reason that I couldn't care less what happened to a rock or tree. And I couldn't care less precisely

because they are not sentient. To see this, consider the metempsychotic version of the original position described earlier, and an exchange between me and my Maker which might go something like this:

God: OK Mark, choose how you want the world to be. I may make you a human, but I may make you a dog. I may even make you a rock.

Me: Fair enough, God. But if I'm a rock, will I be aware of anything. Will I feel, suffer, enjoy, and so on.

God: Of course not: you'll be a rock, for God's sake.

Me: OK then, when I choose how the world is going to be, I don't need to make any provisions for the possibility that you're going to make me into a rock - because if you do then nothing matters anyway.

God: Duh!

The point, of course, is that on the contractarian approach developed here, the scope of morality is restricted to things that an occupant of the original position could rationally worry about being. I can, in the original position, worry about being at least certain sorts of non-human animal since there is something that it is like to be them (at least some of them). That is, non-human animals can, for example, suffer, and if I were one of them I wouldn't want this to happen to me. But if I were told that I was going to be a rock then I couldn't care less what happened to me (and rationally so). Therefore in the original position, I would not vote to include these under the scope of the principles of morality just in case I became one. The contractarian position, then, makes sentience the cut-off point for morality - and does so even though sentience is an undeserved property. And there is no worry of extending the scope of the principle of morality beyond this limit.³⁴

Another objection is raised by Carruthers. According to Carruthers, extension of the contract to non-rational agents would

'destroy the theoretical coherence of Rawlsian contractualism'. He writes:

As Rawls has it, morality is, in fact, a human construction ... Morality is viewed as constructed *by* human beings, in order to facilitate interactions *between* human beings, and in order to make possible a life of co-operative community.³⁵

Actually, I think that Rawls really says nothing of the sort. His theory is completely independent of any story concerning the origin of morality. To suppose that it does is to confuse Kantian and Hobbesian forms of contractarianism. I shall, for the sake of argument, ignore this rather obvious difficulty.

Even if we ignore this, however, the appeal to the origin of morality is rather curious since it automatically leads to a charge of *genetic fallacy*: roughly, the fallacy of confusing the *origin* of morality with the *content* of morality. Carruthers is quick to attempt to head off this charge by emphasizing that he is not claiming that moral statements are really disguised claims about the conditions of survival of the species. This is all very well, but there is, of course, more than one way of committing the genetic fallacy. The basic problem is this. Even if morality were constructed by human beings in order to facilitate interactions between human beings, it does not follow that this sort of origin exhausts the present content of morality, nor that it delimits its scope.

This sort of point is well made by Singer.³⁶ The origin of morality, in fact, might well lie in various mechanisms built in to social animals by a process of natural selection. These mechanisms, in various ways, facilitated social co-operation. If we want to talk about who 'devised' these mechanisms, then the only remotely plausible answer must be *genes*. The mechanisms were 'devised' by genes in order to facilitate the survival of genes through the social co-operation of their gene vehicles (i.e., social animals). Now, I'm not suggesting for a moment that this is true, though it does strike me as substantially more plausible than Carruthers's story. The

point is that even if it is true, it does not follow that the scope and content of morality is restricted to genes, the interaction of genes, and the survival of genes. As Singer points out, morality can develop in ways that are quite distinct from, and even incompatible with, its origin. Therefore, to think that the origin of morality determines the scope and content of morality is to commit the genetic fallacy; and Carruthers has committed this fallacy whatever his protestations to the contrary.

9 Contractarianism and vegetarianism

In developing a contractarian account of animal rights, the hard work lies in defending the claim that non-human animals – and moral patients in general – are recipients of the protection offered by the contract, despite the fact that they are non-rational, and despite the fact that they cannot be regarded as contractors. The task of the preceding sections has been to provide just such a defence. Once this is done, the task of applying the contractarian approach to specific issues involving non-humans is relatively straightforward. Therefore, I do not propose to pursue the question of the application of the contract idea to specific issues in any great depth. Nonetheless, it might be instructive to see how such an application would proceed in the case of one particularly central issue involving non-human animals, namely the moral questions surrounding our raising and killing of non-humans for food.

The original position, it has been argued in this chapter, is a heuristic device that affords us a way of approaching moral problems. In a sense – in the thoroughly heuristic sense explained above – the original position gives us the opportunity to shape – in our minds at least – certain aspects of our world. We can ask ourselves: if I did not know who or what I was, or was going to be, and, therefore, if I did not know what characteristics, powers, aptitudes, needs, and so on I possess, how would I like the world to be? Even in the original position, however, our power to fashion, conceptually

speaking, the ideal world is strictly limited. We can shape the moral and political relations that obtain in our ideal world, but we have no say over its natural order. We have to assume the laws of nature as presently given, and we have to assume that the world is, in all natural respects, the same as the world we presently inhabit. Our brief, provided by the original position, concerns the moral and political order, not the natural one.

According to the contractarian approach, in order to determine the morality of our raising and killing of non-human animals for food – of engaging in animal husbandry in the broadest sense – we have to put ourselves in the original position. In this position, it has been argued, the veil of ignorance excludes knowledge of one's species, and one's status as a rational or moral agent. From the perspective of the original position, the limits of moral considerability, that is, the limits of what one can be morally concerned with, coincide with the class of individuals one can rationally worry about being. And this class can, for present purposes, be subdivided into three: (a) human beings, (b) non-humans typically eaten by humans, and (c) non-humans not typically eaten by humans. In the original position, one does not know into which group one will fall. The two possibilities relevant to the question of the moral status of animal husbandry are, of course, that one might fall into (a) and (b). From the contractarian perspective, the morality of our raising and killing animals for food depends on the rationality of the choices we would make from behind the veil of ignorance. And, to determine this, we have to identify what the members of each category stand to gain and lose from such choices.

Consider, first, what human beings would stand to lose from the widespread adoption of vegetarianism. To choose a world where vegetarianism was morally obligatory for humans would be to choose a world where humans, relative to the actual world, have to give up certain things. The first point to note, however, is that these things do not include life or health. In most environments at least, a healthy life is perfectly possible for a vegetarian, a fact attested to by the existence of millions of healthy vegetarians living

in almost all parts of the world. There is no doubt, of course, that meat is a valuable source of nutrition, primarily because it provides all of the amino acids essential for human beings (i.e., amino acids that the human body is not capable of producing on its own). Meat, however, is not essential in this regard since the essential amino acids can also be obtained from suitable combinations of vegetable protein. And while the knowledge necessary for effecting such combinations is, perhaps, not currently widespread – due in large part to the prevalence of meat-eating in our society – this knowledge is in no way abstruse or recondite. The necessary knowledge is, in fact, no more complex than that required to combine suitable amounts of protein, carbohydrate, and fat in one's diet. Vegetarianism, then, does not ordinarily require humans to give up either their life or health.

If vegetarianism were to become widespread, the principal thing that humans would have to give up would be certain pleasures of the palate. Meat, for most people at least, is often delicious. Some vegetarians actually profess to dislike the taste of meat, others, however, having been persuaded by moral considerations, still dream of rump steaks, and of those heady days when pork ribs would be merrily crackling on the barbecue. While meat is undoubtedly tasty, it is perhaps easy to make too much of this. One who does like the taste of meat dishes is not thereby precluded from finding vegetarian dishes equally appetizing. It is not as if vegetarianism and eating palatable food are mutually exclusive options. To suppose that they are is simply to be ignorant of what it is possible to do with the humble vegetable.

Even if vegetarian dishes are less palatable than meat-based dishes, and it is not clear that they are, we have to weigh up humans' loss of certain pleasures of the palate against what the animals we eat have to give up because of our predilection for meat. Most obviously, of course, they have to give up their lives, and all the opportunities for the pursuing of interests and satisfaction of preferences that go with this. For most of the animals we eat, in fact, death may not be the greatest of evils. They are forced

to live their short lives in appalling and barbaric conditions, and undergo atrocious treatment. Death for many of these animals is a welcome release.

When you compare what human beings would have to 'suffer' should vegetarianism become a widespread practice with what the animals we eat have to suffer given that it is not, then if one were to make a rational and self-interested choice in the original position, it is clear what this choice would be. If one did not know whether one was going to be a human or an animal preyed on by humans, the rational choice would surely be to opt for a world where vegetarianism was a widespread human practice and where, therefore, there was no animal husbandry industry. What one stands to lose as a human is surely inconsequential compared to what one stands to lose as a cow, or pig, or lamb. After all, how many lamb chops would one be prepared to accept for one's life? Therefore, the rational choice must be to opt for a world where vegetarianism was morally obligatory for humans. And if this is the rational choice in the original position, then, if contractarianism is correct, it is the moral choice in the actual world.

If this conclusion is not immediately obvious, then it can be supported by the following considerations. Suppose it were the case that there were two distinct groups of humans. One group, who, following H. G. Wells, we can call the *morlocks*, were cannibals, and they raised and killed another group of humans, the *eloi*, for food, in roughly the same manner in which we now raise and kill non-humans for food. Suppose also that, like us, the morlocks are easily capable of living on vegetable matter alone. The contractarian approach explains why the practice of the morlocks is wrong. In the original position, given that one does not know whether one was going to be a morlock or an eloi, it would be clearly irrational to opt for the above system. If one turned out to be an eloi one would be consigning oneself to a nasty fate, whereas if one turned out to be a morlock one would be perfectly capable of surviving by other means. Therefore, to opt, from behind the veil of ignorance, for the imagined morlock/eloi system would be clearly irrational.

Therefore, if the contractarian approach is correct, this system is immoral.

So, in the case of human beings at least, to opt for the system of human husbandry described above would be clearly irrational. However, once we allow that non-humans are entitled to the protection of the contract, it does not matter whether one is considering a system of human husbandry or a system of animal husbandry. That is, if, in the original position, knowledge of one's species is excluded by the veil of ignorance, then it would be just as irrational to opt for a system of animal husbandry as it would be to opt for a system of human husbandry. Therefore, both systems are, from the standpoint of the original position, equally immoral.

In animal liberationist writings it is common to find the following principle being defended: don't do to animals what you wouldn't be prepared to do to similarly endowed humans. One of the virtues of the contractarian approach, I think, is that it shows clearly the basis of this principle. Once it is allowed that knowledge of one's species should be one of those things excluded by the veil of ignorance, it would be just as irrational to opt for a system that permitted harmful or injurious treatment of non-humans as it would be to opt for a system that permitted the same sort of treatment for humans. From the perspective of the original position, both options are equally irrational. Hence, if the contractarian approach is correct, both are equally immoral. The principle, 'Don't do to animals what you wouldn't be prepared to do to similarly endowed humans' is, then, derivable from the original position once we allow that knowledge of species membership should be bracketed by the veil of ignorance. And given that this is so, the principle can be used as a useful rule of thumb by which to judge our dealings with non-humans.

The above argument, at one point, mentioned the suffering undergone by animals involved in the husbandry industry. This may encourage the thought that if we could somehow eliminate this suffering, the original position might yield a contrary conclusion: that meat-eating is morally acceptable if the animals involved were treated well during their lives and then killed painlessly. On

this suggestion, then, it is not in the killing of animals *per se* that the wrongness of eating meat consists, but in the fact that they are treated so abominably when they are alive. On the present suggestion, the practice of factory farming should be firmly contrasted with that of hunting wild animals for food. This latter practice is morally legitimate, so the argument goes, because the animals here live natural, hence relatively satisfactory, lives. So, according to the present suggestion, if we eliminate factory farming and other intensive rearing practices and eat only non-domesticated or genuinely free-range animals, then our practice of eating meat can be defended.

While it is no doubt true that a world in which the human population ate only wild or genuinely free-range animals would be a morally better world than the actual one, it is still not the case that it would be better than a world where all humans were vegetarians. First of all, the suggestion that we should only eat non-domesticated or free-range animals is completely unfeasible given the present human population. This, however, is not the main problem with the argument. The main problem concerns the morality of the suggestion, not its feasibility. We have the task of comparing two worlds: World 1, where all humans are vegetarians and World 2, where all humans eat only non-domesticated or free-range animals. To judge the relative moral status of the two worlds, one can return to the original position. There is a possibility that you will be a human, and there is also a possibility that you will be an animal killed and eaten by humans. If the former, then vegetarianism requires you to sacrifice certain pleasures of the palate. If the latter, then the failure of humans to adopt vegetarianism requires that you sacrifice your life. A relatively trivial interest of the human has to be weighed against the vital interest of the non-human. Once again there is no contest. If this is not immediately obvious, consider again the situation of the morlocks and the eloi. In H. G. Wells's *The Time Machine*, the eloi were allowed to lead idyllic lives before they were killed. Since they were killed while in a quasi-hypnotic trance, their deaths were also painless. Nonetheless, it would still surely be irrational to opt,

from the original position, for the system of the morlock and eloi. If one turned out to be the latter, one loses one's life. If one turns out to be the former, then one can always survive by other means. Given that it would be irrational to opt for the system of the morlock and eloi, and given that, in the original position one has no knowledge of one's species, it must be equally irrational to opt for a world where animals are eaten by humans, assuming that humans can always survive by other means. And this holds irrespective of the quality of life enjoyed by the animal prior to its being killed and eaten. Therefore, the contractarian position yields the conclusion that vegetarianism is morally obligatory, even if the non-humans we propose to eat are treated well and have happy lives.

Two further consequences of the contractarian position are worthy of note. The first concerns the moral relation between predator and prey. The practice of animal husbandry is often defended, usually more in the popular arena than the philosophical, on the grounds that other animals kill and eat each other. If some animals kill and eat other animals, then why shouldn't we? Or so the argument goes. More precisely, if the lion's preying on the gazelle is not morally wrong, then why should our preying on cows, pigs, or lambs be so? On the other hand, if our preying on cows, pigs, and lambs is morally wrong, then why wouldn't the gazelle have a case against the lion?

This sort of objection is often met with the claim that lions, unlike us, are not moral agents. That is, they are unable to morally assess their actions by dispassionately evaluating them in the light of the moral principles they embody, or with which they conflict. Hence, the lion cannot be charged with doing anything wrong in killing the gazelle. However, this reply, by itself, will not do the work required of it. For, while the lion might not be doing anything for which it can be morally blamed, we might still have a duty to render assistance to the gazelle in precisely the same way in which we might have a duty to prevent a person being harmed by any innocent threat. If a baby has acquired possession of a loaded gun and is firing it in the direction of passers-by, then the baby, not being a moral

agent, is not doing anything for which it can be morally blamed. It is what we can call an innocent threat. Nevertheless, if I am, at fairly minimal risk to myself, able to dispossess the baby, then it seems, *prima facie*, that I have a duty to render assistance to the innocent bystanders by doing so. Similarly, even if the lion cannot be blamed for what it does to the gazelle, we might still have a duty of assistance to the gazelle.

The claim that we have a duty of assistance to prey animals is, as I think will be accepted by most, clearly intolerable. Fortunately, however, the contractarian position does not entail that we have such a duty. In the original position, one of the things one does not know is whether one is going to be incarnated as a predatory animal or as a prey animal, as a carnivore or a herbivore. Given that this is so, to opt for any moral principle which entailed that moral agents have duties of assistance to prey animals would be potentially disastrous. It potentially condemns one to a slow death through starvation. Hence, if one does not know whether one is going to be a carnivore or herbivore, it would be irrational to choose a world which contained such a principle. Even if one turned out to be a herbivore, in fact, the principle would almost certainly prove counterproductive. One of the things one would know in the original position – since one is in possession of all laws pertaining to the natural world – is the role predators play in culling the weaker members of any group of prey animals, thus ensuring the continuing health of the group. One would also know that any situation where a prey animal's natural predator has (almost always at the hand of man) been eradicated is a situation in which the number of prey animals explodes so drastically that disease and starvation are the inevitable result. Even if one turns out to be a prey animal, therefore, opting for a world where moral agents have duties to protect one from one's predators is to opt for a world where the likelihood of one's succumbing to disease or starvation is greatly increased. Thus, there is nothing in the contractarian position which entails that moral agents have duties of assistance to prey animals. With regard to the relation between predators and the

animals upon which they prey, the contractarian position is this: Leave them alone!

A second consequence of the contractarian position concerns certain sorts of human societies living, as we say, on the margins of existence. One important difference between humans and other predators is that humans, typically, are quite capable of surviving without eating meat. This is not, however, true of all humans. There are, of course, certain human societies occupying the margins of existence where eating of meat is essential to survival for the simple reason that supplies of vegetable protein are scarce or non-existent. The *Innuits* provide an obvious example. In such cases, if we assume that the humans involved are unable to relocate to more hospitable climes, then they must be classified as, for all practical purposes, carnivores: creatures unable to survive without eating meat. The original position allows that it is morally acceptable for these people to eat meat. In the original position, to adopt a rule which proscribed this would be irrational; it would be to adopt a rule that potentially sealed one's own fate. More generally, according to the contractarian position, the class of human beings who, for whatever reason, are unable to survive without eating meat should be treated along the same lines as any other carnivore. It is perfectly acceptable for such individuals to eat meat, nor do we have any duties of assistance to the animals upon which they prey.

Perhaps the major argument employed by defenders of the practice of animal husbandry appeals to economic considerations; it focuses on the economic impact that widespread vegetarianism would have, both on those employed in the animal husbandry industry, and on their dependants. The livelihoods of many people – farmers, factory owners, meat packers, transporters, retailers, veterinarians, and so on – are bound up with the success of the animal husbandry industry. Should the industry fail, as it would if vegetarianism became widespread, these people would stand to lose the proverbial shirts off their backs. This would impact not only on those employed in the industry but also on their dependants. Therefore, it can be, and frequently is, argued, the practice of

animal husbandry can successfully be defended by appeal to these sorts of considerations.

Indeed, it might be thought that we could frame these considerations to dovetail nicely with the contractarian framework. In the original position, one does not know what place in society one will occupy. It could turn out, therefore, that one is employed in the animal husbandry industry, or is a dependant of someone so employed. Therefore, it would be irrational to opt for any moral principle which had as a consequence that the animal husbandry industry should be prohibited. In opting for such a principle, one is potentially consigning oneself to financial ruin.

Actually, as I shall try to show, the contractarian position has no truck with this sort of appeal to economic considerations. And any suggestion that it does rests, I think, on a serious misunderstanding of what sorts of considerations can be relevantly incorporated into one's deliberations in the original position. In the remainder of this section, I shall, first, sketch what I think the contractarian position is with regard to the appeal to economic considerations, and, secondly, attempt to diagnose the misunderstanding of the notion of the original position which might lead to the perception that economic considerations are morally relevant.

According to the contractarian position, the appeal to economic considerations is, in the above case, illegitimate. The morality of the practice of raising and killing animals for food should be decided from the original position, and, in that position, the only relevant considerations are what humans and the animals upon which they prey stand to lose from the abandonment or continuation of the practice. But, for reasons which will become clear, economic losses (or gains) cannot legitimately be included as relevant factors in this. The contractarian position, here, can perhaps best be understood in relation to its position on the practice of slavery.

The contractarian position on slavery is that the institution is unjust and, hence, cannot be legitimately defended by appeal to the economic benefits that may accompany it. One cannot legitimately defend an unjust institution by appealing to economic benefits

that it may bring. Thus, the abolition of slavery in, for example, the southern United States was the morally correct course of action, even though it had a devastating impact on the livelihoods of those employed in the cotton industry and on their dependants. In fact, abolition undoubtedly had a calamitous impact on the economy of the South as a whole, leaving the region in a desperate economic slump from which, arguably, it is still recovering today. Nevertheless, according to the contractarian position, since slavery was an unjust institution, abolishing it was the correct thing to do. And it would not be legitimate to attempt to defend slavery by appealing to the desirable economic consequences of its retention. This is not to say that economic considerations can never be morally relevant factors; clearly they can. However, they can be employed as morally relevant considerations only when the institutions they are used to evaluate satisfy certain standards of justice. Slavery does not do so, and therefore cannot be defended by appeal to economic considerations. According to the contractarian position defended here, the same attitude should be taken with regard to the animal husbandry industry. The institution as a whole is unjust, and, therefore, cannot be defended by appeal to economic factors.

The contractarian position in this regard can be made clearer by contrast with that of utilitarianism. All forms of utilitarianism would have to include economic considerations as morally relevant. And, for utilitarianism, whether or not an institution such as slavery counts as just or not is something that can only be decided after all these considerations have been taken into account. For the contractarian, on the other hand, there is a prior standard of justice, and economic considerations will be morally relevant to the evaluation of an institution only if it meets this standard. And the standard is provided by someone in the original position reasoning in accordance with his or her own best interests.

That is the attitude that, I think, the contractarian position takes with regard to the practice of raising and killing non-humans for food, and to any appeal to economic considerations in defence of that practice. What now needs to be shown is exactly *why* the

appeal to economic considerations is, from the contractarian perspective, illegitimate. To see why this is so, consider the following two situations. In Scenario 1, you and three others are sat at a table and you are informed by a fifth person that he is going to distribute £1,000,000 between you. You are further told that one of you will receive £999,999.97 while the other three will receive one penny each. None of you knows, let us suppose, which one will get the fortune. Scenario 2 is importantly different. In Scenario 2, you are again sat at a table with three others and you are informed that you are, collectively, to receive £1,000,000. However, in this scenario, you are required to divide the money into four sums; and prior to this division, you do not know who will receive which sum. The difference between the two cases, then, is that in the former you have no control over the distribution of money; it is something outside your control, a brute fact with which you are simply presented. In the second case, however, the distribution of the money is something which is, so to speak, up for grabs; something to be fixed by negotiation amongst yourselves.

It is important to realize that it is only the second situation that corresponds to the original position. The point of the original position is for you to effect a distribution of, among other things, economic resources and relations, and not to simply be presented with an already existing distribution of such things. So, a situation constituted by (i) an already existing distribution of economic resources and relations, and (ii) a decision that you have to make on the basis of this pre-existing distribution, is not an instance of the original position. The distribution of economic resources and relations is precisely one of the things that is up for grabs in the original position. The claim that economic considerations can legitimately be included in the original position, then, depends on seeing the position as setting the following question: 'Given that economic resources are directed towards the animal husbandry industry in such a way that the individuals employed therein, and their dependants, will suffer serious financial hardship should the industry fail, what moral rules would you like to see adopted?' This,

however, is emphatically not the question the original position sets us. The question of whether economic resources should be directed in such a way is precisely one of the things up for negotiation in the original position. In other words, the economic arguments against vegetarianism would work only if the original position functioned in the manner of Scenario 1; that is, against a background of antecedently assumed economic relations. It does not function in this way, however. The direction of economic resources and the resulting character of economic relations are themselves open to negotiation in the original position.

Once this is understood, it becomes evident that the economic arguments against vegetarianism have no force. In the original position, you are able to opt for a world where there is an animal husbandry industry, and you are also able to opt for a world where there is no animal husbandry industry. The relative moral status of the two worlds depends on the rationality of the choice one makes. Given that you don't know whether you will be human or non-human, eater or eaten, and given that you cannot assume any antecedent economic circumstances, it is fairly clear that to opt for the latter world – the world without an animal husbandry industry – would be the rational choice. And it is this fact, if the contractarian position is correct, which makes the practice of raising and killing animals for food morally wrong.

10 Conclusion

This chapter has been concerned with a contractarian approach to morality in general, and with its application to the concept of animal rights. The approach developed here has been based on the position developed by Rawls, but is, I think, sufficiently distinct from Rawls's position to be dubbed *neo-Rawlsianism*. This neo-Rawlsian position consists, in effect, of a Rawlsian contractarianism that has been purged of its unnecessary Hobbesian elements. The approach developed here is not restricted to the development

of principles of justice, conceived of as governing interactions between individuals and the basic structures of society, but concerns the delineation of principles of morality in general. I have argued that a contractarian approach of this neo-Rawlsian sort can provide a sound theoretical foundation for the attribution of rights to non-human animals. The key to this lies in understanding how two central aspects of Rawls's position converge. The social contract argument depends on the intuitive equality argument in the sense that the latter determines the acceptability of the description of the original position, and this determines which principles of morality we shall accept. Consistent application of the intuitive equality argument, I have claimed, will yield a certain sort of description of the original position. This description of the original position will, in turn, yield principles of morality which apply not only to human beings (i.e., rational agents) but also to many sorts of non-human animals. The final section looked at one application of this general framework: the moral case for vegetarianism.

7

Animal Minds

1 Introduction: morality and mentality

Any defence of the moral claims of animals is going to require that they possess at least some sorts of mental states. If utilitarianism is to work, we are going to have to attribute to animals such states as preferences or desires, or states of pleasure and pain, happiness and unhappiness. If Regan's rights-based approach is to work, we are going to have to regard at least some animals – the ones that make moral claims on us – as subjects-of-a-life, with the mental complexity that this entails. According to the contractarian defence of animal rights, defended in the previous chapter, the limits of moral consideration are determined by what, from the perspective of the original position, one could conceivably worry about being. That is, the boundaries of moral considerability coincide with those of *sentience*. The quickest way to deny animals moral status, therefore, is to deny them mental status; it is to deny that they are the subjects of mental states, to deny that they have a mental life. Such a denial might strike the person on the street – at least if their street is in any way populated with animal life – as absurd. This, however, has not stopped many prominent philosophers from issuing this denial.

It is common to distinguish broadly between two categories of mental states: *sensations* and *propositional attitudes*.¹ Propositional attitudes are mental states that have what is known as *content*. That is, they are states whose ascription to a person involves the use of a

'that'-clause, as in 'Jones believes that the sky is blue'. The class of propositional attitudes includes not only cognitive states like beliefs and thoughts, but also conative and affective states – desiring, hoping, fearing, anticipating, dreading, and so on. Propositional attitudes are not essentially conscious states. One's belief, for example, that Ouagadougou is the capital of Burkina Faso presumably manifests itself only rarely on the conscious stage.² Nonetheless, one can still have this belief even if one is only seldom aware that one has it. Propositional attitudes are what are known as *dispositional* states: whether or not one has a propositional attitude is a matter of one's behaviour, typically one's dispositional behaviour. Thus, for example, if someone asked me if I believed that Ouagadougou was the capital of Burkina Faso, I would reply in the affirmative – assuming I had no desire to deceive or otherwise dissemble. Nor, it is generally accepted, are propositional attitudes defined by a distinctive phenomenology. Believing that one is the owner of a pink Cadillac, for example, is compatible with a variety of different feelings: pride, embarrassment, remorse, and so on. Nevertheless, while propositional attitudes are not essentially conscious states, it is usually thought that they are essentially states that can, in appropriate circumstances, be made conscious. This may sometimes be easy, and if people such as Freud are correct, it may sometimes be extremely difficult. Nonetheless, it is usually thought that propositional attitudes can, at least in principle, be made conscious. And, their capacity for becoming conscious is a property they have essentially; nothing can count as a propositional attitude unless it is, at least potentially, conscious.

While propositional attitudes are, therefore, not unconnected with consciousness, the connection between consciousness and sensations is much tighter. Sensations include bodily feelings like pains, tickles, orgasms, and nausea. States like this do have a distinctive phenomenology; they are defined by the fact that *there is something that it is like* to have or undergo them. Moreover, sensations are not just potentially conscious; they are generally regarded – admittedly with some notable exceptions we shall

examine shortly – as *essentially* conscious. There is, for example, no distinction, between being in pain and feeling pain: to be in pain is just to feel pain and vice versa. If you feel pain at a certain time, and the feeling temporarily stops for a given duration, only to return later then, for the duration of time in which you felt no pain, you were not in pain. And the pain that returns is not the same token pain as the one you initially felt. Rather, it is a distinct token episode of pain.

The category of experiences seems to occupy a curious (and indeed interesting) middle ground between sensations and propositional attitudes. The category incorporates perceptual states such as seeing a pink Cadillac, hearing a loud trumpet, and tasting a sweet strawberry. It also includes quasi-perceptual states such as seeming to see (i.e., hallucinating, or being subject to a visual illusion of) a pink elephant. Whether perceptual or quasi-perceptual, experiences are distinguished from bodily feelings in that they are *about* something; each of them has what is known as an *intentional object*. In this respect, experiences are akin to propositional attitudes. We can talk of the content of both a belief and a perception – and in each case this content is what the state is about. However, experiences are like sensations in that they are also defined by their phenomenology: the character of an experience is defined by the way things seem to its subject when he or she has that experience. Therefore, if we adopt the two-place distinction between sensations and propositional attitudes it is genuinely unclear on which side of the line we place experiences. Do we emphasize their intentionality, and place them with propositional attitudes, or do we accentuate their phenomenological character and group them with sensations?

Thankfully, for our purposes, we do not need to adjudicate this issue. For the purposes of this chapter, two points are crucial. The first is that any defence of the moral claims of animals requires that we be able to attribute to them a mental life of some sort. The second is that there are two distinct avenues available to someone who wished to deny animals a mental life in some or other respect: one

can deny them sensations or one can deny them propositional attitudes. Consider each point in turn.

If you are a hedonistic utilitarian, then inclusion of animals under the scope of moral consideration requires that we attribute to them states of pleasure versus pain, or happiness versus unhappiness. It is genuinely unclear what constitutes happiness, but pleasure is generally regarded as a sensation. Therefore, if we are utilitarians of the basic Benthamite variety, the moral claims of animals requires that they are possessors of sensations. This is captured in Bentham's well-known slogan that the crucial question with regard to animals is not whether they can reason but whether they can *suffer*. Singer, on the other hand, is a preference utilitarian. Therefore, defending the moral claims of animals is going to require showing that they are the possessors of preferences. And preference is a propositional attitude. In fact, preference is a certain type of desire: a desire that a particular alternative or situation should obtain as opposed to others. Regan's case for animal rights was predicated on the assumption that the claim that certain animals – at the very least all mammals – are subjects-of-a-life, where an individual is a subject-of-a-life only if it has beliefs, desires, perceptions, memories, and so on, and a sense of the future, including its own future. These are all propositional attitudes. Therefore, Regan's case for animal rights seems to require the ascription of propositional attitudes to animals. There are good reasons for thinking that the contractarian defence of animal rights developed in the previous chapter also requires the attribution of propositional attitudes to animals. From the perspective of the original position, in order to decide how one would like the moral and social world to be if one were a member of species S, one must have some idea of the *preferences* of typical members of S. Moreover, it is plausible to suppose that possession of desires by an organism entails that it also possesses beliefs. It is difficult to see, for example, how one can desire a book (or, more precisely, desire that one have the book) without believing that one does not presently have the book. Possession of a certain desire entails possession of certain associated beliefs.

And, so, if the contractarian defence requires attribution of desires to non-humans, it also requires attribution of beliefs. The same would be true of the positions developed by Singer and Regan.

Consider, now, two avenues available to someone who wished to deny that animals have a mental life in some or other respect. She could deny that animals have sensations, or she could deny that they have propositional attitudes; or, of course, both. Each denial however, is not of equal status. If you deny that animals have even sensations, then it is difficult to avoid denying that they have propositional attitudes. A propositional attitude is the sort of state that must be at least capable of becoming conscious. A creature with no sensations would be a creature devoid of consciousness; and therefore not the sort of creature capable of possessing propositional attitudes. To deny a creature sensations is thereby to deny them propositional attitudes. However, the converse is not true. It is possible for a creature to possess bodily feelings without having achieved the level of psychological complexity required for propositional attitudes.

If we put these two points together, we arrive at the following conclusion. Any case for the moral claims of animals is going to require that we at least be able to attribute to them sensations – bodily feelings and their ilk. If we can't even do this, there is no case for including animals under the umbrella of moral consideration. Therefore, defending the claim that animals are conscious in the basic sense of being possessors of sensations is a *sine qua non* of the moral case for animals. However, from the point of view of developing the moral case of animals, it would be better if we could also demonstrate that they possess propositional attitudes as well. Failing this, it may well be that the moral case for animals would rest solely upon hedonistic utilitarianism. And this would be an unsound and extremely restricted basis for this case.

In this chapter, I am going to argue that many animals can be regarded as subjects of both sensations and propositional attitudes – at least, there is no reason for supposing that they are not such subjects and every reason for thinking that they are. To this

end, I shall examine, and I hope demolish, the standard objections of philosophers to regarding animals in this way.

2 HOT and the rejection of animal consciousness

The most direct way of blocking the inclusion of non-humans under the moral umbrella would be by denying that they have a mental life. At one time, such a denial was widely accepted, and inspired by such notable philosophers as Descartes and Malebranche. In present times, however, any blanket denial of the mental life of animals strikes most intelligent laypersons – at least those with any familiarity with animals – as patently absurd. This, however, has not stopped some philosophers continuing the tradition of Descartes and Malebranche. Indeed, I think it would be true to say that philosophy has provided a consistently hostile stance towards the claims of animals to be mental subjects – at least in some or other respect. This hostility has typically taken one of two forms. On the one hand, there is hostility to the idea that animals can be regarded as possessors of propositional attitudes. This form typically accepts that animals are conscious, but denies that they have the psychological complexity to be regarded as bearers of propositional attitudes. On the other, there is hostility to the idea that animals can even be conscious. I am going to begin by looking at the second sort of hostility.

Hostility to the idea that animals are even conscious derives from *higher-order thought* (HOT) models of consciousness. Such models have been developed by Rosenthal (1986, 1990, 1993) and Carruthers (1996, 1998) among others. According to such models, very roughly, a mental state *M* is conscious if, and only if, the subject of that state possesses a thought about *M*: a thought to the effect that she has *M*. One perceived implication of this is that if you are not capable of having a higher-order thought about your mental states, then those states are not conscious. This implication has typically been used as a way of attacking HOT models (e.g., Dretske 1995).

Anything that does not have concepts of mental states will not be able to have a thought about those states. Therefore, HOT models rule out the attribution of consciousness to anything that does not have a *theory of mind*. This exclusion will encompass not only animals but also young children – a conclusion that many will regard as frankly incredible. Rosenthal, the most influential defender of HOT attempts to distance himself from this conclusion; arguing that, appearances notwithstanding, HOT accounts do not have this implication. Carruthers, however, embraces it. Animals and young children do not, in fact, have conscious mental states.

I have argued elsewhere that HOT accounts face several serious logical problems.³ Here I shall focus on just one of these: the problem of *regress*. Therefore, any worry that they rule out the attribution of conscious states to animals (and young children) is not to be taken seriously. This section will be concerned with explaining exactly what HOT accounts are (and what they are not). Subsequent sections will develop the case against these accounts.

In order to properly understand HOT models of consciousness, we first need two distinctions: (i) the distinction between *creature* and *state* consciousness, and (ii) the distinction between *transitive* and *intransitive* consciousness.

Creature versus state consciousness. We can ascribe consciousness both to *creatures* (e.g., John is conscious as opposed to asleep or sedated) and to mental *states* (my belief that Ouagadougou is the capital of Burkina Faso is, at this point in time, a conscious belief). HOT accounts are attempts to explain state consciousness not creature consciousness.

Transitive versus intransitive consciousness. We sometimes speak of our being conscious *of* something (e.g., of how the sun dances on the bright, blue water). This is *transitive* consciousness. Transitive consciousness is a form of creature consciousness. Mental states are not conscious of anything. Rather, creatures are conscious of something in virtue of the mental states they have. *Intransitive* consciousness, on the other hand, can be ascribed

both to creatures and states. A creature is intransitively conscious when it is conscious as opposed to asleep, knocked out, or otherwise unconscious. A state is intransitively conscious when we are consciously entertaining it – as when I think to myself that Ouagadougou is the capital of Burkina Faso.

The core idea of HOT models is that intransitive state consciousness can be explained in terms of transitive creature consciousness. This idea can be divided into two claims. First, (and roughly), a mental state *M*, possessed by creature *C*, is intransitively conscious if and only if *C* is transitively conscious of *M*. Second, a creature, *C*, is transitively conscious of mental state *M* if and only if *C* has a thought to the effect that it has *M*. Thus, intransitive state consciousness is to be explained in terms of transitive creature consciousness, and transitive creature consciousness (at least in this context) is to be explained in terms of a higher-order thought – a thought about a mental state.

Within this general framework, HOT models can be developed in two different ways which, following Carruthers, we can label *actualist* and *dispositionalist*.⁴ According to actualist accounts, a mental state *M* possessed by a creature *C* is intransitively conscious if *C* has an occurrent thought to the effect that it possesses *M*. An occurrent thought is one that is currently being entertained by the subject (although, as we shall see, not necessarily *consciously* entertained). According to *dispositionalist* forms of HOT, on the other hand, the presence of an occurrent thought about state *M* is not necessary for the intransitive state consciousness of *M*. All that is required is that the creature *C* be disposed to have a thought to the effect that it possesses *M*.

3 The problem of regress

HOT models try to ground the intransitive consciousness of mental state *M* in a given subject, *C*, by the simultaneous possession by *C* of a thought to the effect that they have or are in *M*. For

example, my pain is conscious if and only if I possess a thought to the effect that I am in pain. However, an obvious dilemma arises. Either the higher-order thought to the effect that I am in pain is itself intransitively conscious or it is not. If it is, and if its being intransitively conscious is what grounds the intransitive consciousness of the pain, then the HOT model clearly faces a regress. The property that confers intransitive state consciousness has not been identified but deferred. This is the first horn of the dilemma. To avoid this, the HOT account is committed to the claim that the higher-order thought that confers intransitive consciousness on a mental state is not, itself, intransitively conscious.

This, however, leads directly to the second horn of the dilemma. Suppose that the higher-order thought that (allegedly) confers intransitive consciousness on my pain is not itself intransitively conscious. Then, among other things, I will not be aware of thinking that I am in pain; I will, in effect, have no idea that I am thinking this. But how can this thinking that I am in pain make me aware of my pain if I have no idea that I am thinking that I am in pain?

Take my belief that Ouagadougou is the capital of Burkina Faso. When I entertain this belief, it makes me aware of a fact – the fact that Ouagadougou is the capital of Burkina Faso. However, it is only rarely that I entertain this belief: most of the time this belief is one of my unconscious mental states. What makes it unconscious? Well, precisely that it does not make me aware of the fact that Ouagadougou is the capital of Burkina Faso. I still possess the belief because I am disposed to become aware of the fact under certain eliciting conditions (e.g., someone asks me what is the capital of Burkina Faso). However, when it is in unconscious form, as it is most of the time, it is unconscious precisely to the extent that it does not make me conscious of anything. Therefore, the HOT account faces the rather pressing and difficult problem of explaining how a non-conscious thought can make us conscious of anything at all.

The HOT theorist can object that the higher-order thought that confers intransitive consciousness on my pain is an *occurrent* thought, and this makes it crucially different from my belief that

Ouagadougou is the capital of Burkina Faso, which does not make me aware of the relevant fact because it exists in *dispositional* form. So what we must try and understand is: (i) how we can entertain a thought in a non-conscious manner, and (ii) where doing so makes us aware of what the thought is about. How can a non-conscious thought that we are currently entertaining make us aware of what that thought is about?

We can, I think, make sense of the idea of currently entertaining an intransitively non-conscious thought. Suppose I think unconsciously – perhaps due to various mechanisms of repression – the thought that someone very close to me is seriously ill. What would this mean? We might explain it in terms of various unexplained feelings of melancholy that assail me when I am talking to them, or a vague sense of foreboding that I can't quite pin down or give shape to. The thought is occurrent because it is playing an active role in shaping my psychological life. The truth of this account can be contested, but it does at least make sense. However, what we can make no sense of is the idea that this unconscious but occurrent thought makes us aware of what it is about. If it were to do this, then I would have to be aware of the fact that my friend is seriously ill. But as soon as I am aware of this fact, then I am consciously thinking that my friend is seriously ill. That is, the thought has become intransitively conscious. To undergo the dawning realization that my friend is seriously ill is precisely to think – to consciously think – that my friend is seriously ill. That is precisely in what the awareness of the fact of my friend's dire health consists.

Therefore, while we might be able to make sense of the idea of occurrently entertaining a non-conscious thought, what we cannot make sense of is the idea that this thought should make us aware of what it is about. That it does not make us aware of what it is about is precisely what it is for the thought to be non-conscious. Conversely, as soon as it does make us aware of what it is about, it is a conscious thought – because making us aware of what it is about is precisely what it is for the thought to be conscious.

The moral is clear. Intransitively non-conscious thoughts do not make us transitively conscious of what they are about. But the HOT account tries to explain the intransitive conscious of a mental state – say my pain – in terms of our being transitively conscious of that state. But making me transitively aware of my pain is precisely what an intransitively non-conscious higher-order thought cannot do. If it does, it is thereby an intransitively conscious higher-order thought. And if it is an intransitively conscious higher-order thought, we are back to the problem of regress.

Dispositionalist versions of HOT, of the sort defended by Carruthers, might seem to avoid this version of the regress problem, but only by way of a gambit that is truly desperate, and a resulting view that is grossly implausible. According to dispositionalist accounts, intransitive state consciousness of a mental state M consists in the subject of M – creature C – being *disposed* to have a thought to the effect that it possesses M. The intransitive consciousness of my pain consists in the fact that I am disposed – I have a tendency – to form a thought to the effect that I have it.

This account seems seriously misguided on at least two counts. First, intransitively conscious experiences or sensations seem categorical in a way that dispositions are not. My pain, for example, is something that is actually taking place within me. It does not seem the sort of thing that could be constituted by a disposition or tendency for something to take place in me. Similarly, the conscious experience of hearing a loud trumpet seems to be either something I am having or not having at any given time. And this does not seem to be the sort of thing that can be captured in terms of dispositional facts about myself. On the face of it, there seems almost as much difficulty in understanding how the intransitive consciousness of a sensation or experience could consist in a disposition to produce a higher-order thought as by, say, a disposition to behave in various ways. Phenomenologically, conscious experiences and sensations are categorical in a way that makes them incapable of being constituted by dispositions.

Second, there seem to be severe difficulties in even making sense of the claim that dispositions can actually *make* something

be a certain way. Dispositions are simply not the sorts of things that can make a given item have a property P. They are descriptions of that item's tendency to acquire P under certain conditions. It is not the brittleness of an item that makes it break upon falling. Rather, its brittleness is simply its tendency to break in these sorts of circumstances. So it is, to say the least, difficult to see how a disposition to instantiate a higher-order thought could even be the right sort of thing to make a mental state intransitively conscious.

Dispositionalist forms of HOT are grossly implausible. Actualist forms of HOT have a problem of regress. Therefore, even if HOT could be used to undermine the idea that animals are conscious, this is not something that should engender any sleepless nights for the defenders of the moral claims of animals.

4 The holism of the mental

The case against the possibility of attributing propositional attitudes to non-humans has been developed by several authors, most notably Donald Davidson and Stephen Stich. Part of Davidson's argument coincides, in all important respects, with that of Stich, and I shall, therefore, discuss them together. The conceptual centrepiece of this argument is a principle sometimes referred to as the *holism of the mental*.

Suppose, following Norman Malcolm, we watch a dog chase a cat that runs up an oak tree, and then disappears from sight.⁵ The dog remains barking at the foot of the tree, looks upwards, and so on. It seems plausible to suppose that the dog believes that the cat is up the tree (and that the dog desires to catch the cat – the belief and desire combination collectively explaining the dog's behaviour). However, Davidson argues that this supposition is problematic. He writes:

Can the dog believe of an object that it is a tree? This would seem impossible unless we suppose the dog has many general beliefs about trees: that they are growing things, that they need soil and

water, that they have leaves or needles, that they burn. There is no fixed list of things someone with the concept of a tree must believe, but without many general beliefs there would be no reason to identify a belief as a belief about a tree, much less an oak tree. Similar considerations apply to the dog's supposed thinking about the cat.⁶

The moral, according to Davidson, is this:

We identify thoughts, distinguish between them, describe them for what they are, only as they can be located within a dense network of related beliefs. If we really can intelligibly ascribe single beliefs to a dog, we must be able to imagine how we would decide whether the dog has many other beliefs of the kind necessary for making sense of the first.⁷

But this creates a problem:

It seems to me that no matter where we start, we very soon come to beliefs such that we have no idea at all how to tell whether a dog has them, and yet such that, without them, our confident first attribution looks shaky.⁸

The principle assumed in these passages is the principle of the *holism of the mental* which, following Davidson, I shall understand as primarily a principle governing the *attribution* of beliefs (and other propositional attitudes) to individuals. The principle can be stated as follows:

Holism of the mental. The attribution of a single belief or other propositional attitude to an individual requires, and only makes sense in terms of, the attribution of a network or system of related beliefs (or other propositional attitudes).⁹

As the first passage makes clear, Davidson's worries in this regard ultimately stem from concerns about the possibility of attributing content to the supposed belief of the dog. This is reflected in the fact

that the principle of the holism of the mental, a principle governing the attribution of beliefs to individuals, derives from a distinct thesis, one which concerns the identity of the content, or meaning, that beliefs possess. We can refer to this latter thesis as the principle of *content holism*:

Content holism. The content of a belief (or other propositional attitude) possessed by an individual is determined by the relations that content bears to the contents of all other beliefs (or other propositional attitudes) possessed by that individual.¹⁰

Thus, we cannot attribute content to the alleged belief of the dog (i.e., that the cat is up the tree) because content is fixed by the content of other beliefs; such as the claim that trees are growing things, that they need soil and water, and so on. Therefore, we cannot attribute the requisite content to the dog since the surrounding beliefs that are constitutive of that content are missing. Content, however, is essential to any belief. A belief is individuated by way of its content. The belief that the cat is in the tree, for example, is distinguished from the belief that the cat is in the house by way of the distinct contents of the beliefs. Therefore, since content is essential to beliefs, if we cannot attribute the content, we cannot attribute the belief. Therefore, we cannot attribute to the dog the belief that the cat is up the tree.

A similar line of argument is developed by Stich. Stich also sees the crucial difficulty in attributing beliefs to non-humans as deriving from a problem in identifying their content. In 'Do Animals Have Beliefs?' Stich argues that the question of what content can be ascribed to non-humans is moot.¹¹ The question has no answer. In his later book, *From Folk Psychology to Cognitive Science*, he takes a somewhat more conciliatory stance when he claims that the question is hopelessly *context-relative*. In some conversational contexts, ascription of content = based states to a non-human animal would be correct, but in other contexts, ascription of the same content-based state to the same animal at the same time would be incorrect.¹²

Stich's argument is based on his analysis of the concept of belief, and the corresponding notion of belief content. According to Stich, the relation of content-identity is, in fact, a similarity relation. The notion of the content of a belief can be factored into three elements: causal-pattern similarity, ideological similarity, and reference similarity.

Two beliefs counts as similar along the dimension of *causal-pattern similarity* if they have similar patterns of potential causal interaction with other beliefs and with (actual or possible) stimuli. In addition to *global* causal-pattern similarity, there are various dimensions along which a pair of beliefs can be *partially* causal-pattern similar. For example, a pair of beliefs may interact similarly with other beliefs in inference, but may have different links with stimuli. These beliefs would count as similar when the context focuses on inferential connections, but as rather dissimilar when the context focuses interest on the connections between belief and perception. Causal-pattern similarity is the feature that is focused upon by classical functionalist accounts of content.

The second sort of feature used to assess similarity of beliefs is what Stich calls *ideological similarity*. The ideological similarity of a pair of beliefs is a function of the extent to which the beliefs are embedded in similar networks of belief. Ideological similarity measures the 'doxastic neighbourhood' in which a given pair of belief states find themselves. As in the case of causal-pattern similarity, partial ideological similarity is often more important than global ideological similarity. Since belief states are compound entities, ideological similarity can be assessed separately for the several concepts that compose a belief. And context can determine which concepts are salient in the situation at hand. For example, suppose the context focuses on 'bourgeois'. Then if Jack and Jill both say, 'Abstract art is bourgeois', we may count them as having similar beliefs if their other beliefs invoking the bourgeois concept are similar, even if they have notably different beliefs invoking their abstract art concept. But if the difference in their conceptions of abstract art looms large in the context, our

judgement will be reversed, and they will not count as having similar beliefs.

The third sort of feature used in assessing belief state similarity is *reference similarity*. According to Stich, two beliefs count as reference similar if the *terms* the subjects use to express the beliefs are identical in reference. What actually fixes reference is not an easy matter to decide. One prime candidate is the causal history of the term, a causal chain stretching back through the user's concept, through the concept of the person from whom he acquired the term, and so on back to the person or stuff denoted. A second candidate, defended by Wittgenstein, Burge, and others, is the use of the term in the speaker's linguistic community. Neither of these accounts is free from difficulties. And Stich does not wish to adjudicate between these accounts. He does say, however, that in his view context is an important determinant of reference.

According to Stich, therefore, the notion of sameness of content, hence the notion of sameness of belief, is a complex concept which straddles all three features of causal-pattern, ideological, and reference similarity. Depending on the context of discussion, one or more of these factors can assume primary importance.

When we attribute content to an individual, either human or non-human, therefore, our attribution is carried by one or more of these three factors. Stich, however, claims that reference similarity is inapplicable to non-humans. The reason is that, according to Stich, a pair of beliefs count as reference similar if the *terms* the subjects use to express the beliefs are identical in point of reference. This characterization, of course, automatically makes the concept of reference similarity inapplicable to non-language using creatures. Therefore, reference similarity can play no role in determining the content of the beliefs of non-linguistic creatures. Determination of the content of such beliefs, therefore, is solely a matter of the causal pattern and ideological network in which the beliefs are embedded. But this entails that there are strongly holistic constraints on the concept of content-identity. If causal-pattern and ideological similarity are the only determinants of

the content of a belief, then such content will be solely a function of the causal and ideological relations that the belief bears to other beliefs. However, it is very implausible to suppose that the causal-pattern and ideological networks of beliefs present in non-human animals will be in any way similar to those present in humans. And, therefore, Stich claims, since this is all we have to go on in the case of non-humans, we are unable to attribute content to non-human animals. But, since content is essential to beliefs and other propositional attitudes, this means that we are unable to attribute these states to them also. The prospect of attributing propositional attitudes to non-humans founders on the impossibility of attributing content to them.

Therefore, Stich, like Davidson, identifies the problem of attributing beliefs and other propositional attitudes to non-humans as deriving from the problem of attributing content to such creatures. And, in both cases, the problem in attributing content to them, derives from the holism, and the constraints it imposes, that is constitutive of content. In fact, in both Davidson and Stich, we find essentially the following argument:

- (1) We can attribute a belief (or other propositional attitude) to a non-human animal only if we can attribute content to that belief.
- (2) We can attribute content to the belief of a non-human animal only if it possesses a broad network of related beliefs that is largely similar to our own.
- (3) Non-human animals do not possess a broad network of beliefs that is largely similar to our own.
- (4) Therefore, we cannot attribute content to the beliefs of non-human animals.
- (5) Therefore, we cannot attribute beliefs to non-human animals.

The same argument applies, *mutatis mutandis*, to all propositional attitudes. This is by far the most important philosophical

argument against attributing mental states to animals. Indeed, I think that ultimately it is the only remotely plausible argument against doing so. Fortunately, at least for the moral claims of animals, the argument fails. The remainder of the chapter will be concerned with explaining exactly why this is so.

5 An unsuccessful refutation

At this point it may be worthwhile to pause and consider Tom Regan's attempt to meet the challenge provided by the holism of the mental.¹³ As I shall try to show, Regan's challenge fails, but its failure is, I think, nonetheless instructive.

Regan begins by distinguishing two conceptions of what is involved in possessing a concept of a given object *x* (e.g., a tree). The first of these is what he calls the *all-or-nothing view*. According to this view, the concept of *x*, possessed by an individual, is constituted by *all* the beliefs held by that individual regarding *x*. Thus, if two individuals differ with regard to the beliefs they hold about *x*, then they have distinct concepts of *x*. Thus, since the beliefs I have about trees differ from the beliefs had by our imagined dog, we possess distinct concepts of a tree. Thus, it is not possible to attribute to the dog the belief that the cat is up the tree. The dog does not possess the concept of a tree and, therefore, cannot possess any beliefs about trees.

As Regan points out, there are serious problems with the all-or-nothing view. It is implausible to suppose that any of us share precisely the same beliefs about anything. And, if the all-or-nothing view were correct, this fact would preclude us from sharing any concepts. If every belief we might have about trees is relevant to the determination of the concept of a tree, then, it seems overwhelmingly likely that each one of us, or at least most of us, will have distinct concepts of a tree. But this would render communication about trees impossible. If the all-or-nothing view of concepts were correct, it is not communication, but equivocation, which would

be the rule. And the possibility of attributing beliefs to anyone, not only dogs, would be undermined.

Regan contrasts the all-or-nothing view with what he calls the *more-or-less* view. According to this view, concept possession is a matter of degree. The more beliefs two individuals share about *x*, the more they share the same concept of *x*. Children, for example, who know that trees are growing things, that they need soil and water, but not that they produce the energy they require by photosynthesis share, to some extent, our concept of a tree. The more-or-less view has distinct advantages over the all-or-nothing view. Primarily, it allows us to make sense of the attribution of concepts, and hence beliefs, to normal adult members of our own culture (something the all-or-nothing view was hard pressed to do) and also to children, to members of different cultures, to the mentally enfeebled, and so on (something the all-or-nothing view could certainly not do).

Assuming that the more-or-less view of concepts is preferable to the all-or-nothing view, Regan then attempts to use it to establish the validity of attributing content to non-humans. In this context, Regan introduces the idea of a *preference-belief*. Let us call our imagined dog 'Brenin', and give him a bone (perhaps to make up for the disappointment of missing out on the cat). Can Brenin have the belief that we are giving him a bone? Given the arguments of Davidson and Stich, this will depend on whether Brenin possesses the concept of a bone, and this, in turn, will depend on whether he has the requisite beliefs about bones. But, as Regan points out, we can be reasonably sure that Brenin possesses at least one belief about bones: he believes that bones satisfy certain desires and are to be chosen to satisfy those desires. That is, he believes that the bone will satisfy his desire for a particular flavour and should be chosen if he wishes to satisfy that desire.

Thus, Brenin's behaviour surely licenses attribution to him of at least this preference-belief. But, Regan argues, this is one of the beliefs that define the content of our concept of a bone. Therefore, given the more-or-less view of content possession, we can say that Brenin possesses our concept of a bone, at least more-or-less.

This attempt to enfranchise Brenin is certainly ingenious. Unfortunately, it too possesses serious problems. Firstly, it is not clear that the switch to the more-or-less view solves the problem of communication afflicting the all-or-nothing view. Vastly different sorts of information can, on the more-or-less view be associated with possession of the same concept. But if we assume that it is information that gets transmitted in communication, the problem of equivocation still looms large. Secondly, it is at least arguable that the plausibility of the more-or-less view stems from its confusing two quite different senses of 'more-or-less believing that p'. On the one hand there is the relatively innocuous idea that agents can differ in their *epistemic commitment* to p (I will nail my flag to p; you grant p only your provisional assent). This idea is not at issue. On the other hand, there is the idea that *content identity* is a matter of degree. There is a big difference between the claim that one can more or less believe that p, and the claim that what one believes is more or less p. And it is the second claim that is at issue here. The idea that there might be something that is almost, but not quite, the proposition that there is a bone buried in the yard seems to make little sense.

Third, and most importantly, however, even though Davidson and Stich clearly reject the all-or-nothing view of concept possession, they would still not accept Regan's premise that possession of a single belief is sufficient for possession of a concept, even to a degree. Davidson, for example, writes:

There is no fixed list of things someone with the concept of a tree must believe, but without many general beliefs there would be no reason to identify a belief as a belief about a tree, much less an oak tree.¹⁴

Possession of a concept does not require possession of any fixed list of beliefs; hence the all-or-nothing view should be rejected. However, it does require a certain threshold number of beliefs. Possession of a single preference belief about a bone, for example, is presumably not sufficient for possession of the concept of a bone.

What Regan has presented us with, in effect, is a false dilemma. We are not forced to choose between an all-or-nothing view of concept possession which makes communication between distinct individuals a practical impossibility, and a more-or-less view which entails that possession of even a single belief is sufficient for possession of a concept. Davidson and Stich adopt a third alternative: possession of many beliefs about *x* is required for possession of the concept of *x*, but possession of any fixed list of beliefs is not. We might call this a *cluster theory* of concept possession.

Therefore, I think we should conclude that Regan's attempt to psychologically enfranchise animals fails. At this point we may encounter a temptation to which many defenders of non-human mentality have succumbed. As DeGrazia, for example, points out, there is no straightforward inference from the claim that the content of animal beliefs is inexpressible by us to the claim that, therefore, they possess no content.¹⁵ There is no reason to suppose, that is, that content must be expressible by us humans in order to count as content. This claim is quite correct, but tempts us into yielding too much to the argument from holism. The temptation is to simply yield and allow that while non-humans may possess beliefs and other propositional attitudes, the content of those attitudes is forever inexpressible by us. They have beliefs, but we can never know what those beliefs are. I think that this grants far too much to the arguments of Davidson and Stich. The notion of content is ambiguous, and in one sense of 'content' it is, in fact, perfectly possible to make accurate and determinate attributions of content, and hence of content-based states, to non-human animals. In next section, I shall try to show why.

6 Attributing content

The argument to be developed in this section runs as follows. First, I shall argue that the content of any belief is constituted by two factors. On the one hand, there is the *mode of presentation* of the

belief, that is, the way or manner in which the belief represents what is known as its intentional object, that is, the thing that the belief is about. On the other, there is the *reference* of the belief; the object that the belief takes as its intentional object. The content of the belief supervenes on, or is determined by, both factors taken together. Secondly, and in consequence, the attribution of a belief to an individual can be carried by, or based on, either of these factors. Sometimes, our attribution of a belief to an individual will be based on the way or manner in which that belief represents its intentional object; sometimes it will be carried by the intentional object itself and not its mode of representation. Thus, attribution of a belief can, in different contexts, be sensitive to distinct and non-reducible factors, and this affects our conception of the beliefs thus attributed. Third, attributions of belief based on, or sensitive to, the referent of the belief, and not its mode of representation, are not constrained by the sorts of holistic considerations adduced by Davidson and Stich. That is, an attribution of a belief to an individual that is sensitive to the object of the belief, and not the mode of representation of that object, does not depend on, or in any way require, that the belief be embedded in a network of beliefs. Fourth, and finally, attribution of beliefs to non-humans can therefore be carried, and justified, on the basis of knowledge of the reference of those beliefs; knowledge of the mode of representation of the beliefs is not necessary.

Content and reference

The content of any belief (or other propositional attitude) is constituted by two distinct and non-reducible factors. The content of any belief depends on its intentional object, that is, on the object the belief is a belief about. Part of the content here derives from what is known as the *mode of presentation* of the object; that is, the way in which the object is represented. But part of the content derives from the object itself, and not from its mode of presentation. This point was brought to prominence by the now classic thought experiments

developed by Hilary Putnam and Tyler Burge.¹⁶ Here is Putnam's version. The basic idea is now well known, so I shall be brief.

We are to conceive of a near duplicate of our planet: twin-earth. Not only are the physical environments of earth and twin-earth largely identical, but many inhabitants of Earth have duplicate counterparts on twin-earth. These counterparts are type-identical with their corresponding earthlings with respect to physical constitution, and also with respect to experiential and dispositional histories, where these are specified non-intentionally. The key difference between the two planets is that the liquid on twin-earth that runs in rivers and taps, although qualitatively identical with the liquid that we, on earth, refer to with the term 'water', is not in fact water, but, rather, a distinct substance. Thus, although the twin earthlings refer to this substance with the term 'water', it is not water since it is not a substance with a chemical structure made up of two parts of hydrogen to one part of oxygen but, instead, has a complex chemical structure - XYZ. Let us call this substance *retaw*. Now, if Herbert₁ is an English speaker of Earth, and Herbert₂ is his twin-earth counterpart, then it is fairly clear that when Herbert₁ says 'water is wet', and Herbert₂ produces an utterance of the same phonetic form, they say something different. Their utterances have distinct meanings. And this is true even though, *ex hypothesi*, what is going on in their heads is identical. Meanings, as Putnam points out, are not in the head. Moreover, these differences go on to effect oblique occurrences of those sentences that specify the content of the respective Herbert's beliefs. Thus, Herbert₁ believes that water is wet. But Herbert₂ cannot have this belief. Herbert₂ believes that retaw is wet (even though he would express this belief by an utterance of the form 'water is wet'). Thus, the content of the two Herberts's beliefs differ, and thus their beliefs differ, even though what is going on in their heads is the same.

Putnam's thought experiment works by driving a wedge between the two factors that determine the content of belief. The beliefs of Herbert₁ and Herbert₂ have the same mode of

representation of their objects. That is, they represent water and retaw respectively in precisely the same way – as colourless, odourless, drinkable, et cetera liquids. Nonetheless, the contents of the beliefs, and hence the beliefs themselves, differ. And this shows that the content of a belief cannot be entirely determined by its mode of representation of an object. Content is, in part, constituted by the referential properties of beliefs.

We are now in a position to see just what a crucial – and entirely unjustified – move was Stich's refusal to apply the concept of reference similarity to non-human animals. Stich, remember, thinks that the notion of content-identity is a similarity relation which can be factored into three components: causal-pattern similarity, ideological similarity, and reference similarity. But Stich also thinks that the relation of reference similarity is inapplicable to non-human animals: since they have no language, reference similarity is out of the question. This follows from his characterization of reference similarity as a relation holding between the *terms* or *expressions* of a language and the world. That is, one of the relata of the relation of reference is always a linguistic entity. However, Stich gives no justification or defence of this characterization. Stich does allow that some sort of derivative reference relation might obtain between the representations possessed by non-human animals and the world. But he does not regard this as important enough to warrant the inclusion of reference similarity as a determinant of the content of animal belief states. In my view, this is to get the order of primacy reversed. Reference is a relation which holds *primarily* between internal states of creatures and the world, or between the behaviour of creatures and the world, and *derivatively* between terms or expressions and the world. And this claim will be defended later in the chapter. For now, it is sufficient to point out that in excluding the relation of reference from the factors involved in attributing beliefs to non-humans, Stich is excluding an important constituent of the content of beliefs. No wonder he regards the attribution of beliefs to non-humans as problematic.

Reference and transparency

The fact that the content of any belief depends on two distinct factors means that an attribution of a belief to an individual can be made on the basis of either factor. Our attributions of belief, that is, answer to two distinct interests we have in such attributions. When the context, for example, calls for us to be interested in the way in which an individual represents an object, our attribution will typically be carried by, and hence sensitive to, the mode of representation of that object. Thus, if, for example, we were interested in the respective behaviours of the two Herberts, how they interact with water or retaw, whether they will drink it, wash in it, and so on, then our attribution of belief to them would be based on, and carried by, the mode of representation of water or retaw. Other contexts, however, might require sensitivity to the referential constituent of the content. Questions of truth and falsity, for example, require sensitivity to this latter constituent.

Putnam's example, in effect, shows that there are two distinct ways of looking at, or individuating, the content of any belief, hence two distinct ways of looking at, or individuating, that belief. Contents of beliefs, hence beliefs themselves, can be individuated *narrowly* or *broadly*. A belief narrowly individuated comprises the mode of representation of the intentional object of that belief. It is the belief narrowly individuated that plays a causal-explanatory role in the agent's psychology. That is, it is not the relation to the referent that is causally or explanatorily relevant to the agent's behaviour, but the way in which that referent is represented internally. It is not *what* is represented that matters to causal or explanatory role but *the way in which* it is represented. On the other hand, it seems undeniable that beliefs have representational or semantic properties – for they have truth-conditions – and these properties exert a pull in the individuation of beliefs by content. The belief individuated by way of its referential or semantic properties is a belief broadly individuated. Putnam's thought experiment thus shows that these two ways of individuating beliefs are not equivalent

and, in certain circumstances, can come apart. Thus, the two Herberts are identical in point of their beliefs narrowly individuated, but distinct in point of their beliefs broadly individuated. And, therefore, we must recognize that belief content is essentially a hybrid of conceptually disparate elements, *both* of which inform our conception of belief and its individuation.¹⁷

Because the same belief can, in different contexts, be subject to distinct standards of individuation, our attribution of beliefs and other propositional attitudes to individuals is fundamentally ambiguous. In certain cases, the attribution will be carried by the mode of representation of the object of the belief. In certain other contexts, however, the attribution will be carried by the referential properties of the belief; that is, it will be based on the intentional or represented object of the belief, and not the manner in which this object is represented. There is no question of which attribution is the correct one; both are equally legitimate. And which sort of attribution we make is, in any particular case, largely a function of our interests underlying this attribution.

There is, in fact, a familiar linguistic device we use to record the fact that distinct standards of individuation can be applicable to, and hence govern the attribution of, the same belief. There are what is known as *transparent* attributions of a belief, and there are *opaque* attributions of that belief. Equivalently, there are what is known as *de re* attributions of a belief and *de dicto* attributions. An opaque or *de dicto* attribution of a belief is an attribution of a belief individuated narrowly, a belief individuated by way of its mode of representation of its object. A transparent or *de re* attribution of a belief is an attribution of a belief individuated broadly, individuated by way of its semantic or referential properties. Suppose I am in a room with many people, one of whom is Jones. I believe that Jones is a spy, and thus the opaque, *de dicto*, attribution of that belief to me would be correct. However, unbeknownst to me, Jones is also the tallest man in the room. There, given I believe that Jones is a spy, and given that Jones is also the tallest man in

the room, the transparent or *de re* attribution to me of the belief that the tallest man in the room is a spy would also be correct. This is true even if I vehemently deny the attribution – because, for example, I mistakenly believe that Smith is the tallest man in the room. Similarly, put in these terms, in Putnam’s twin-earth case, the two Herberts are subject to the same opaque or *de dicto* attributions of belief but distinct transparent or *de re* attributions of belief.

There is a linguistic device we use to flag the difference between opaque/*de dicto* and transparent/*de re* attributions: opaque/*de dicto* attributions are made using a *that*-clause; transparent/*de re* are made using an *of*-clause. We say: I believe *that* Jones is a spy; but we say I believe, *of* the tallest man in the room, that he is a spy. Both opaque/*de dicto* and transparent/*de re* attributions of belief are equally legitimate, but they are driven by different interests, governed by distinct standards of individuation for beliefs, and appropriate in different contexts. Amongst philosophers, in recent decades, there has been a pronounced tendency to privilege opaque or *de dicto* attributions of belief over transparent or *de re* ones. Transparent/*de re* attributions are regarded as suspicious or in some way defective. The reasons for this stem largely from Quine, and worries he adduced concerning the legitimacy of so-called *quantifying in*. These do not concern us. For purposes, it is enough to note that this privileging of opaque/*de dicto* over transparent/*de re* is not part of common sense. We switch comfortably between the two types of attribution, we have a familiar linguistic device that allows us to do so without ambiguity.

Transparency and holism

Suppose an explorer comes across a hitherto undiscovered primitive tribe. The tribesmen and women, let us suppose, are terrified of his camera. Since they have never seen a camera before, and since they are manifestly terrified of it, it is clear that their concept of a

camera occupies a place in a vastly different causal and ideological network of beliefs than our own. Nevertheless, it would surely be misguided to insist that therefore, they can possess no beliefs or other propositional attitudes towards the camera. And it would be equally misguided to claim that it is not possible to specify the content of their beliefs. Indeed, adequately explaining their behaviour would require both postulation of belief and of a determinate content. Thus, for example, we might explain their terror by hypothesizing that they believe the camera will take their soul. Unless one was in the grip of a theory, one would surely allow that the tribe's persons can bear this sort of belief about the camera.

The procedure we would employ to make sense of their behaviour would presumably begin with a transparent or *de re* attribution of belief. Puzzled by their behaviour, we first need to identify the object that is eliciting it. We need to make sure that it is the camera, and not, say, our hats (which we always wear when taking photographs of them) that prompts this behaviour. Once we do this, we say, with perfect legitimacy, that they have some strange beliefs *of or concerning* the camera. Once we have done this, we can then investigate precisely how they must be thinking of the camera in order to behave in the way they do. Now we have switched from the *de re* project of identifying the *object* that elicits their strange behaviour to identifying the *mode of presentation* of the object – the way they are representing the camera – that explains this behaviour.

If we were influenced by the Davidson/Stich argument, we might rescind from making opaque attributions of belief to them. We might not allow ourselves the luxury of saying the tribesmen and women believe *that* the camera will steal their souls. But even if we are willing to do this, there is no justification for rescinding from the corresponding transparent attribution: they believe, *of* the camera, that it will steal their souls. And once we have identified the relevant ways in which the tribesmen and women represent the camera to themselves – once we have identified the beliefs they hold about the camera – there is no justification for refusing

to make opaque attributions – as long as we accept that the content they ascribe is distinct from the content ascribed in attributions of camera beliefs to us.

A transparent attribution of the belief to a tribesman does not require that the tribesman possesses a concept of the camera whose content is individuated by its place in a network of beliefs. The tribesman, let us suppose, does possess a network of beliefs about the camera, and this network is very different from our own. As a result, he possesses a concept, narrowly individuated, of a camera that is very different from our own. However, the transparent attribution of beliefs about the camera is in no way affected by this fact. The transparent attribution of the belief, in this case, is sensitive to, and carried by, the camera itself, and not the mode of presentation of the camera. It is only when the content of a belief is individuated narrowly, by its mode of presentation, that attribution of a belief is subject to the sort of holistic constraints that concern Davidson and Stich. When the content of a belief is individuated broadly, in terms of the beliefs object itself and not its mode of presentation, then attribution of this belief is not subject to these constraints. And this is why it is possible to legitimately attribute beliefs and other content bearing states to individuals who possess vastly different causal and ideological networks than our own – even, crucially, when we are unaware of the structure of those networks.

The arguments of Davidson and Stich, then, even if correct, can have application only to situations in which the attribution of beliefs and other propositional attitudes is opaque. Transparent attributions of beliefs are not affected by the sorts of holistic constraints identified by Davidson and Stich.

Attributing content to animals

Precisely the same principles apply in the case of non-human animals. A dog's concept of a tree, it is plausible to suppose, occupies

a place in a vastly different system of beliefs and other propositional attitudes than does our corresponding concept. The dog, presumably, does not know that trees are growing things, that they require soil and water, that they drop leaves and needles, that they burn, and so on. And this, it is argued, is grounds for denying the dog the concept of a tree. Hence, it is grounds for denying the dog any beliefs about trees, including the belief that the cat is up the tree.

This argument, however, will work only in the case of opaque/*de dicto* attributions of belief. In such attribution, belief content is individuated narrowly, in terms of the mode of presentation of the intentional object of the belief. Such attributions are, therefore, sensitive to, and carried by, the mode of presentation of the belief's object, since an object's mode of representation in a belief does depend on other beliefs possessed by a subject. Transparent/*de re* ascriptions, on the other hand, depend on individuating belief content broadly, by reference to the intentional object of the belief itself and not that object's mode of presentation. Since the identity of the belief's intentional object is not determined by its mode of presentation, such attributions are not in any way dependent on this mode of presentation. And, therefore, transparent attributions of belief are not subject to the sort of holistic constraints that govern opaque attributions.

Therefore, the fact that a dog lacks many, or even all, of the beliefs about trees that humans typically possess can, at most, impact on the possibility of making opaque attributions of belief to him. Transparent attributions, on the other hand, are not affected by this deficit. Therefore, to argue that the dog cannot have the belief that there is a cat up the tree because it would be impossible to specify the content of any such alleged belief ultimately rests on a failure to adequately recognize that there are two distinct ways of individuating the content of a belief, hence two distinct ways of individuating beliefs, and therefore, two distinct ways in which beliefs can be attributed.

This defence of attributing beliefs and other propositional attitudes to animals is anticipated by Davidson. He writes:

Someone may suggest that the position occupied by the expression 'that oak tree' in the sentence 'The dog thinks the cat went up that oak tree' is, in Quine's terminology, transparent. The right way to put the dog's belief (the suggestion continues) is 'The dog thinks, with respect to that oak tree, that the cat went up it' or 'That oak tree is the one the dog thinks the cat went up.'¹⁸

According to Davidson, however, the problem with this suggestion is that:

But such constructions, while they may relieve the attributer of the need to produce a description of the object that the believer would accept, nevertheless imply that there is some such description; the *de re* description picks out an object the believer could somehow pick out.¹⁹

This is true. *De re* ascriptions of belief, at least arguably, ultimately presuppose *de dicto* ascriptions in this sense: to believe something *of* an object requires that one have some way of *picking out* that object. One cannot simply think of an object, one must, necessarily, always think of an object in some way or other – one must think of the object under a mode of presentation. What is puzzling, however, is why Davidson thinks this presents any problem for the possibility of making *de re* ascriptions of belief to animals. Why should we not simply accept that (i) the dog thinks, *of* the oak tree, that the cat ran up it, and (ii) there is a mode of presentation of the tree that enables the dog to grasp that tree in thought and belief? We must be careful to distinguish two importantly distinct claims: (i) the possibility of making a transparent attribution of a belief to an individual presupposes that there is a mode of presentation of the object of that belief that the individual grasps, and (ii) the possibility of making a transparent attribution of a belief to

an individual depends on our knowing what this mode of presentation is. Claim (ii) does not follow from (i), and there is little reason for thinking that (ii) is true. Certainly, nothing so far identified in the Davidson/Stich argument permits us to infer (ii). It is here that we arrive at the crucial assumption operative in Davidson's argument:

I assume that an observer can under favourable circumstances tell what beliefs, desires and intentions an agent has. Indeed, I appealed to this assumption when I urged that if a creature cannot speak, it is unclear that intensionality [i.e. opacity] can be maintained in the descriptions of its purported beliefs and other attitudes.²⁰

Here, as Wittgenstein would put it, is the decisive movement in the conjuring trick. This assumption, on its own, precludes the attribution of beliefs to animals. If we cannot tell, even in favourable circumstances (whatever they are) what beliefs, desires, and intentions an agent has, then we have no business thinking that it is the proper subject of beliefs, desires, and intentions. But why should we assume this? The assumption is by no means obvious. Indeed it seems to be an expression of a form of verificationism that slides from what we can know to the way things have to be – a verificationism that has been thoroughly discredited in other contexts. Whether or not the charge of verificationism can be sustained, what is most important is that *this is not an argument, it is an assumption*. Accordingly, Davidson has not argued that animals cannot have beliefs; he has assumed that they cannot.

Instead of making this assumption, why not simply accept – as the analogy with the exotic tribe was intended to show – that, with respect to animals we can, with relative confidence, make transparent or *de re* attributions of belief, while also acknowledging that opaque or *de dicto* attributions are far more problematic. This is not to say that we can never, in principle, make opaque attributions. Rather, it is to acknowledge that such attributions cannot

be made from the armchair, but have to be based upon detailed, perhaps painstaking, observation of animal behaviour. To give a general flavour of the sorts of empirical considerations that would be relevant, let's return to the dog that chases, say, a squirrel up the tree.

If the dog is to believe that the squirrel ran up the tree, then it must have beliefs about squirrels. And to have beliefs about squirrels it must have the concept of a squirrel. Does it have the concept of a squirrel? I take no stand – as I said this is not the sort of thing to be judged from the armchair. But consider the sorts of things we would have to ask (and answer) in order to successfully adjudicate this matter. For example, we would have to decide whether the dog can successfully discriminate squirrels from other small mammals. If it cannot, then it does not have the concept of a squirrel, but of something else. Perhaps what the dog possesses is what, following Gibson (1979), we might call an *affordance*-based concept: it has the concept of a *chaseable* thing – under which it subsumes squirrels, rabbits, rats, and other small mammals. If so, then the mode of presentation under which the dog thinks about the squirrel is *chaseable thing* – and we can then attribute to the dog the *de dicto* or opaque belief that the chaseable thing ran up the tree (assuming, of course, we could similarly identify the relevant mode of presentation of the tree). Or perhaps the dog is an experienced hunter, and has come to differentiate between things that try to escape by running up trees (squirrels, cats), and things that try to escape by going down holes (rabbits, rats). This, after all, can impact significantly on the hunting strategy he adopts. If so, the mode of presentation under which the dog thinks of the squirrel might be: chaseable thing that can go up trees; and we can attribute to the dog the *de dicto* or opaque belief that the chaseable thing that can go up trees in fact went up a tree.

My point is not that these are correct *de dicto* attributions, are correct ways of understanding beliefs of the dog. Rather, it is that for every legitimate transparent attribution of a belief to a dog, there is every reason for supposing that there is a correct opaque

attribution. Moreover, there is no reason for supposing that this opaque transformation is something that is necessarily closed to us. Opaque attributions may be problematic. They may require extensive ethological or behavioural research. But there is no reason for supposing that we can never make opaque attributions of belief to animals.

To summarize: we can make transparent attributions of belief to animals. There are opaque attributions that can be made – even if we don't know, in particular cases, what they are. And these claims, together, are all we need to defend the claim that animals are the subjects of beliefs and other propositional attitudes.

7 Belief and truth

There is a way of reading Davidson according to which the argument I have just presented – based on the holism of the mental – is merely a preliminary or softening up argument. The real argument, some think, turns on the relation between belief and *truth*. Davidson writes:

First: I argue that in order to have a belief, it is necessary to have the concept of belief. Second, I argue that in order to have the concept of belief one must have language.²¹

The second premise is relatively uncontroversial. Accordingly, most of the work Davidson needs to do consists in defending the first premise. Why does having a belief require having the concept of belief? On the face of it, this seems to implausibly over-intellectualize the idea of belief. Davidson's defence of this turns on the idea of *surprise*:

Surprise requires that I be aware of a contrast between what I did believe and what I come to believe. Such awareness, however, is a belief about a belief: if I am surprised, then among other things I come to believe that my original belief was false.²²

However, this in turn, seems to over-intellectualize the concept of surprise. Consider a recent experimental finding. If you play a dog a video recording of a man's face apparently speaking, but with a woman's voice superimposed over his voice, then a dog will stare longer at this recording than in the normal case where a man's face goes with a man's voice. The converse point applies when a man's voice is superimposed over a woman's face. The dog, it seems reasonably secure to suppose, recognizes that *something is up*. Could we say that the dog is surprised? Not on Davidson's characterization of surprise. For that involves the claim that to be surprised one has to be able to think to oneself that one has a belief and this belief is false. This seems an artificially inflated understanding of surprise. We could make the same move in connection with the dog's realization that *something is up*. To believe that something is up requires one to think to oneself that one has a belief and this belief is problematic in some way. This interpretation of thinking that something is up is, frankly, ridiculous. So, why should we not say the same thing about Davidson's understanding of the concept of surprise?

The answer is that Davidson is not really concerned with the idea of surprise at all, but with a particular ability: the ability to entertain one's beliefs and ascertain their truth or falsity. One cannot have beliefs at all without this ability, because to have beliefs is to have belief about the correctness of one's beliefs:

What I do want to claim is that one cannot have a general stock of beliefs of the sort necessary for having any beliefs at all without being subject to surprises that involve beliefs about the correctness of one's own beliefs. Surprise about some things is a necessary and sufficient condition of thought in general. This concludes the first part of my 'argument'.²³

Then the question, however, is why should we accept this? Davidson *asserts* that we cannot have a general stock of beliefs without having beliefs about the correctness of one's beliefs, but gives no argument for this. He seems to realize that he doesn't have an argument

for this point – that is why he puts the term ‘argument’ within scare quotes.

Overall, his ‘argument’ seems to be this:

- P1. One can have a single belief only if one has a general stock of beliefs (the holism of the mental).
- P2. One can have a general stock of beliefs only if one has beliefs about one’s beliefs.
- 3. Therefore, no creature incapable of having a belief about a belief is capable of belief.

Premise 1 is dubious – at least in the form Davidson wants to understand it. And, indeed, the idea of the holism of the mental is a lot less fashionable than it used to be.²⁴ Premise 2 is nothing more than an undefended assertion. There is little here, I think, over which the defender of animal mentality need lose any sleep.

8 How animals can refer

I have argued that, in arguments over whether it is possible to ascribe propositional attitudes to non-human animals, we must be careful to observe the distinction between opaque and transparent attributions. An opaque attribution of a propositional attitude to an individual depends essentially on the mode of presentation of the intentional object of the attitude; that is, on the way in which that object is represented to the individual. A transparent attribution, on the other hand, depends only on the identity of the intentional object, and not on the way in which that object is represented. Opaque attributions, being dependent on the mode of representation of intentional objects, are crucially dependent on the causal and ideological networks of beliefs in which the attributed attitude is embedded, since these are partial determinants of the mode of presentation of the attitude’s intentional object. Transparent attributions, however, are not similarly dependent on such networks of

attitudes, since it is not generally true that the mode of presentation of an intentional object determines the identity of that object. Therefore, while opaque attributions of propositional attitudes to non-humans might be undermined by the fact that non-human minds are constituted by vastly different networks of beliefs than our own, transparent attributions are not undermined by this fact. The arguments of Davidson and Stich, then, can, at most, jeopardize opaque attributions of propositional attitudes to non-humans; they in no way undermine transparent attributions of such attitudes.

We are now in a position to see just what a crucial move was Stich's refusal to apply the concept of reference similarity to non-human animals. Stich, remember, thinks that the notion of content-identity is a similarity relation which can be factored into three components: causal-pattern similarity, ideological similarity, and reference similarity. But Stich also thinks that the relation of reference similarity is inapplicable to non-human animals. This follows from his characterization of reference similarity as a relation holding between the terms of a language and the world. However, Stich gives no justification or defence of this characterization. Stich does allow that some sort of derivative reference relation might obtain between the representations possessed by non-human animals and the world. But he does not regard this as important enough to warrant the inclusion of reference similarity as a determinant of the content of animal beliefs and other propositional attitudes. Stich, however, is definitely swimming against the current on this point. The orthodox view is that reference is a relation which holds *primarily* between internal states of individuals and the world and only *derivatively* between terms or expressions and the world. And this is, ultimately, why it is possible to make transparent attributions of propositional attitudes to individuals, both human and otherwise. The remainder of this chapter will be concerned with explaining how the internal representations of non-human animals might be able to refer to, or represent, the world.

One of the principal projects of recent philosophy of mind has been the attempt to provide an account of the relation of

representation. This has been particularly important insofar as the notion of representation has been seen as the basis of the relation of intentionality. The intentionality of mental states, at least according to one prominent account, derives from the representational relations holding between internal states and the world. The account of representation I shall now consider is by no means the only possible account; and it does have its opponents. However, it is, in my view, the best philosophical account currently available. It is worth noting that even if the following should prove an inadequate account of representation, all competing accounts construe representation as a natural relation holding primarily between internal states of individuals and the world. Thus, even if the following account turns out to be wrong, all competing accounts are compatible with the claim that the notion of reference is applicable to non-language using animals. Thus, they are all compatible with the practice of making transparent attributions of prepositional attitudes to animals. The account I favour has been developed most fully by Ruth Millikan, and is known as the *teleological theory*.²⁵

A teleological theory of mental representation will employ, as a pivotal concept, what Millikan calls *relational proper function*. The proper function of some mechanism, trait, or process is what it is *supposed* to do, what it has been *designed* to do, what it *ought* to do. The concept of proper function is a normative concept. Proper functions can come about either through the intentions of a designer, or through a mindless process such as natural selection. A hammer has the proper function of knocking in nails in virtue of the intentions of its designers, makers, and users. A heart, on the other hand, has the proper function of pumping blood in virtue of various pressures of natural selection.

The proper function of an item is defined in terms of what that item *should* do, not what it actually does or is disposed to do. The concept of proper function, being normative, cannot be defined causally or dispositionally. What something does, or is disposed to do, is not always what it is supposed to do. To use a flagship example of Millikan's, the proper function of a sperm cell is to

fertilize the ovum. The vast majority of sperm cells, however, do not accomplish this task. The proper function of an item is its *Normal* function, where, following Millikan, the capitalized 'N' indicates that this is a normative sense of normal as opposed to a causal or dispositional one.

The Normal function of many evolved items is *relational* in character. In the case of evolved organisms, function is ultimately relative to gene reproduction; that is, the function of many evolved characteristics is ultimately to enhance reproductive capacity. Generally this means that the characteristic is to enable the organism (the gene vehicle) to cope with its environment: to locate food, evade predators, protect itself against heat and cold, and so on. It is here that the relationality of Normal or proper function arises. Normal functions are often defined relatively to some environmental object or feature: the function of the chameleon's skin is to make the chameleon the same colour as its immediate environment; the function of the lion's curved claws is to catch and hold on to prey; the function of the bee's dance is to help other bees locate nectar, and so on. In each case, the function of the characteristic is specified in terms of a relation to an environmental item. And the reason for this is that the very reason the characteristic in question exists is that it has evolved to meet certain environmental pressures.

The core idea of the Ideological theory of mental representation is that the mechanisms responsible for mental representations are evolutionary products also. As such, they will have relational proper functions. The idea, then, is that the representational features of a given cognitive mechanism derive from the environmental objects, properties, or relations that are incorporated into that mechanism's relational proper function. That is, if a cognitive mechanism M has evolved in order to detect environmental feature E, then this is what makes an appropriate state S of M about E; this is what gives the state S the content that E. In this way, the representational content of cognitive state S derives from the relational proper function of mechanism M that produces S.

This account requires a clear distinction to be drawn between a cognitive state and a cognitive mechanism. Roughly, the distinction will be implemented in the following way. An organism's cognitive state tokens are (often) caused by events occurring in that organism's environment. And there are mechanisms, typically neuronal, that mediate those causal transactions. Each of these mechanisms will, presumably, have an evolutionary history and, therefore, will possess a proper function. And, it is plausible to suppose, this proper function will be precisely to mediate the tokenings of cognitive states. That is, on the teleological view, there are various neural mechanisms whose proper or normal function is to produce tokenings of cognitive states in environmentally appropriate circumstances. According to the teleological theory, the content of these cognitive states derives from the environmental features that are incorporated into the proper functions of the mechanisms that produce these states. Thus, if cognitive state *S* is produced by mechanism *M*, and if the proper function of *M* is to produce *S* in environmental circumstances *E*, then, according to the teleological theory, *S* represents, or is about, *E*. In this way, the content of a cognitive state derives from the relational proper function of the mechanism that produces it.

There is nothing in this story that requires cognitive states – beliefs, desires, and so on – to themselves have proper functions. It is perfectly consistent to claim that the content of a cognitive state derives from relational proper function while denying that the cognitive state itself has that proper function. And this is good for the teleological account, because the claim that cognitive states such as beliefs and desires have proper functions is notoriously difficult to defend.

The teleological theory does not purport to provide a complete theory of *content*. If it did, it would attract the obvious objection that mechanisms and structures of organisms can have relational proper functions and yet not have propositional content. It does not seem appropriate, for example, to assign semantic content to

hearts, despite their relational proper function. Rather, the teleological theory is advanced as a theory of the referential component of representation. It does not try to explain why a state might represent an organism in a particular way. It does try to explain why a state can represent, in the sense of refer to, denote, or pick out, a particular object. The teleological theory only purports to be a theory of a part of content: that part of content that is constituted by the referential component of the representational relation.

This being so, the teleological theory is best viewed within the framework of the two-factor account of propositional attitudes developed earlier. As we have seen, our intuitive conception of content is constituted by two distinct factors. On the one hand there is a mode of presentation of the intentional object of a belief; on the other, there is the intentional object itself. The content of the belief is constituted both by the mode of presentation of the object, and by the representational relation the belief bears to the object itself, independently of its mode of presentation. Consequently, there are two distinct ways in which a propositional attitude might be attributed: opaquely or transparently. Transparent attributions of belief and other propositional attitudes do not require detailed knowledge of the mode of presentation of the intentional objects of belief; they merely require that we know *which* objects are the intentional objects of beliefs. Transparent attributions of beliefs to non-human animals, therefore, will be legitimate as long as we have some reason for thinking that non-human animals have internal states that function to represent, or pick out, environmental (or bodily) items. This is where the Ideological theory comes in. If we understand representation in the way suggested by the teleological theory, then it becomes essentially a biological phenomenon. Internal states can represent environmental items in virtue of the fact that the mechanisms which produce such states have evolved to produce them in circumstances where a given environmental item is present. The representational properties of the state derive from the evolutionarily determined relational proper function of the biological mechanisms that produce them. Representation, therefore,

is ultimately a biological notion. And, given that non-human animals clearly do have internal mechanisms which have evolved to detect certain environmental features, it is perfectly appropriate, at least in principle, to make transparent attributions of beliefs and other propositional attitudes to them.

Therefore, for example, when you arrive home and your dog on the inside scratches at the door while you, on the other side, are fumbling for your key, it is perfectly legitimate to ascribe to the dog the belief that you are on the other side of the door. We don't have to know how the dog represents you, that is we don't have to know under what mode of presentation the dog subsumes you. And presumably the dog's mode of presentation of you is radically different from that mode of presentation whereby you represent yourself to yourself, or whereby you are represented by others. All we need to know, in order to make the transparent attribution of this belief to him, is that he detects *you*. What initially grounds our confidence that he detects you, as opposed to anyone else, is the difference in his behaviour when another person approaches the door (the tone and cadence of the bark changes, and so on). And what justifies our confidence that there is some sort of detection going on here is, ultimately, evolutionary theory. We know, on evolutionary grounds, that dogs are going to have evolved mechanisms to detect friends from foes, familiar from strange animals, pack members from outsiders. Thus, if the teleological theory is true, we know on evolutionary grounds that some sort of representation is going on here. And we know from careful observation of behaviour, what the object of the representation is. And this is all we need to know in order to make a transparent attribution of belief.

9 Conclusion

Philosophical worries about attributing mental states to non-humans have two principal sources. First, there is the higher-order thought model of consciousness which, if correct, *might*

undermine the attribution of conscious states to non-human animals. Fortunately, this model is not correct. Secondly, there are the worries raised by Davidson and Stich turning on the holism of the mental and the relation between belief and beliefs about beliefs. Fortunately, to the extent that we find arguments here, they are not convincing. And the rest is simply assertion masquerading as argument.

Here is a parting question. One of the things about philosophy that has always struck me as curious is a peculiar sort of blindness philosophers seem to bring to bear on their discussions of animals. When they talk about animals, good philosophers, even great ones, seem to make the sorts of mistakes they wouldn't make in other contexts, and so manage to convince themselves of the most outlandish of views. The question is: why is this? For example, when Stich, a very good philosopher indeed, blithely assumes that reference similarity is a relation that holds between language and the world, rather than the mind and the world, thus precluding this relation from characterizing the beliefs of non-humans, why does he do this? The mistake, I am certain, is not in any way wilful. Why does Davidson – arguably one of the great philosophers of the twentieth century, and a master constructor of arguments – seem to rely so heavily on an 'argument' for denying beliefs to animals that even he recognizes is stunningly weak: an argument that in other contexts he would, I'm pretty sure, never have accepted? Again, this is not, I am certain, in any way wilful. And how could *anyone* ever have convinced themselves that animals are not even conscious?

Descartes, the Father of modern philosophy, famously managed to convince himself that animals are automata. And in this respect the Father never seemed to quite lose his grip on the children. Why this should be is a topic for another time.

Notes

1 Animal rights and moral theories

1. Peter Singer, *Animal Liberation* (New York: The New York Review of Books 1975). Reprinted by Thorsons (1991). All page references are to the latter.
2. Peter Singer, *Practical Ethics* (Cambridge: Cambridge University Press 1980); 'Utilitarianism and vegetarianism', *Philosophy and Public Affairs* 9, 8 (1980); 'Animals and the value of life', in *Matters of Life and Death*, ed. T. Regan (New York: Random House 1980); 'Killing humans and killing animals', *Inquiry* 22 (1979); 'All animals are equal', *Philosophical Exchange* 1, 5 (1974).
3. Tom Regan, *The Case for Animal Rights* (Berkeley: The University of California Press 1984). Reprinted by Routledge (1988). All page references are to the latter.
4. The following have also been influential: Stephen Clark, *The Moral Status of Animals* (Oxford: Oxford University Press 1977); Mary Midgley, *Animals and Why They Matter* (Harmondsworth: Penguin 1984); S. F. Sapontzis, *Morals, Reasons, and Animals* (Philadelphia: Temple University Press 1987); James Rachels, *Created From Animals: The Moral Implications of Darwinism* (Oxford: Oxford University Press 1990); David deGrazia, *Taking Animals Seriously* (Cambridge: Cambridge University Press 1996).
5. Rosalind Hursthouse, *On Virtue Ethics* (Oxford: Oxford University Press); Michael Slote, *Morals From Motives* (Oxford: Oxford University Press 2001). For a development of virtue ethical thinking particularly relevant to the case of animals, see Rosalind Hursthouse, *Ethics, Humans and Other Animals* (New York: Routledge 2000).
6. Peter Carruthers *The Animals Issue: Moral Theory In Practice* (Cambridge: Cambridge University Press 1992).

7. Regan, *The Case for Animal Rights*, pp. 163–74.
8. I refer to my *Animals Like Us* (London: Verso 2002).
9. I borrow the terminology of *Hobbesian* and *Kantian* contractarianism from Kymlicka's 'Contractarianism', in P. Singer ed., *A Companion to Ethics* (Oxford: Basil Blackwell 1989).
10. John Rawls, *A Theory of Justice* (Oxford: Oxford University Press 1971). However, as I shall argue in Chapter 6, Rawls's version of contractarianism is vitiated by several crucial unexpurgated Hobbesian assumptions.
11. For some reasons on this score, see my *The Philosopher and the Wolf* (London: Granta 2008).

2 Arguing for one's species

1. A somewhat similar scenario is to be found in Colin McGinn's example of the vampires who are capable of living on orange juice. See his *Moral Literacy, or How to do the Right Thing* (Cambridge, MA: Hackett 1992), pp. 21–2.

3 Utilitarianism and animals: Peter Singer's case for animal liberation

1. Singer, *Animal*, pp. 2–23.
2. See especially, Singer, 'Utilitarianism and Vegetarianism'.
3. J. J. C. Smart, 'An outline of a system of utilitarian ethics', in J. Smart and B. Williams eds, *Utilitarianism: For and Against* (Cambridge: Cambridge University Press 1973), pp. 18–21.
4. I sometimes think that philosophers, being necessarily a cerebral bunch, often tend to be somewhat out of touch with those more somatically inclined. It seems to me, having conducted a completely unscientific survey of some of my more hedonistically oriented friends, that many people would in fact volunteer to be hooked up to the machine for life. Hence I do not necessarily endorse this criticism of utilitarianism. I merely record it.
5. Smart, in fact, describes the machine in this more general way. I have distinguished the pleasure machine from the more general experience machine simply for expository purposes.

6. Rawls, *A Theory of Justice*, p. 24.
7. Regan, *The Case for Animal Rights*, pp. 208–11.
8. G. E. Moore, *Ethics* (Oxford: Oxford University Press 1912).
9. Rawls, *A Theory of Justice*, pp. 23ff.
10. Regan, *The Case for Animal Rights*, pp. 208ff.
11. J. S. Mill *Utilitarianism, Liberty, Representative Government* ed. A. D. Lindsay (London: J. M. Dent & Sons 1968); R. M. Hare, 'Rights, utility and universalization: Reply to J. L. Mackie', in R. G. Frey ed., *Utility and Rights* (Minneapolis: University of Minnesota Press 1984); J. Harsanyi, *Essays on Ethics, Social Behaviour and Scientific Explanation* (Dordrecht: D. Reidel 1976); J. Griffin, *Well Being: Its Meaning, Measurement and Moral Importance* (Oxford: Oxford University Press 1986).
12. R. M. Hare, 'Rights, utility and universalization: A reply to J. L. Mackie', p. 106.
13. Rawls, *A Theory of Justice*, pp. 31, 450, 564.
14. Singer, 'All animals are equal', p. 155.
15. Regan, *The Case for Animal Rights*, pp. 230ff.

4 Tom Regan: animal rights as natural rights

1. Regan, *The Case for Animal Rights*, p. 243.
2. Regan regards utilitarianism as an essentially teleological theory, in the sense explained in the previous chapter.
3. Regan, *The Case for Animal Rights*, pp. 326–7.
4. *Ibid.*, pp. 267–73.
5. *Ibid.*, p. 272.
6. *Ibid.*, pp. 284–5.
7. *Ibid.*, p. 305.
8. *Ibid.*, p. 298.
9. *Ibid.*, p. 308.
10. *Ibid.*, p. 331.
11. *Ibid.*, p. 333.
12. It is worth noting that it is not possible to derive the worse-off principle from the contractarian position I shall develop in Chapter 6. The contractarian position does not license a blanket 'save the ones made worse-off' claim. What also has to be factored into contractarian considerations is the likelihood of oneself being one of the

worse-off or one of the better-off. This, I think, is a strength of the contractarian position. For a more extended discussion of this point in connection with the harm of death, see my *Animals Like Us* (London: Verso), ch. 5.

5 Virtue ethics and animals

1. Elizabeth Anscombe, 'Modern moral philosophy', *Philosophy* 33 (1958): 1-19. Perhaps the most influential recent development of virtue ethics is Rosalind Hursthouse, *On Virtue Ethics* (Oxford: Oxford University Press 1999).
2. Roger Scruton, *Animal Rights and Wrongs* (London: Demos 1996).
3. Rosalind Hursthouse *Ethics, Humans and Other Animals* (London: Routledge 2000).
4. *Ibid.*, p. 159.
5. *Ibid.*
6. *Ibid.*, p. 161.
7. *Ibid.*
8. *Ibid.*, pp. 161-2.
9. I would like to thank my colleague Brad Cokelet for suggesting this option to me.
10. He also says it is about fun. But as we have seen, virtue trumps fun.
11. Milan Kundera *L'Insoutenable Légèreté de L'Être* (Paris: Gallimard 1983), p. 76. Translation is mine.

6 Contractarianism and animal rights

1. Immanuel Kant, 'Duties to animals and spirits', in his *Lectures on Ethics* trans. L. Infield (New York: Harper & Row 1963).
2. Peter Carruthers, *The Animals Issue* (Cambridge: Cambridge University Press 1992), pp. 96-8.
3. Tom Regan, 'The case for animal rights', in P. Singer ed., *In Defence of Animals* (Oxford: Basil Blackwell 1985), p. 17.
4. For a much more balanced discussion, see David deGrazia, *Taking Animals Seriously* (New York: Cambridge University Press 1996), pp. 166-210.

5. As I mentioned in Chapter 1, the work of Will Kymlicka provides an important exception to this claim. I am not sure, however, if Kymlicka will agree with the way I am going to develop the distinction.
6. I have chosen the 'Hobbesian' versus 'Kantian' terminology in deference to Kymlicka, who has done more than anyone to make this distinction clear. One of the disadvantages of this terminology, however, is that on it even Kant emerges as not a consistent Kantian contractarian.
7. This has also been recognized by Martha Nussbaum in *Frontiers of Justice* (Cambridge, MA: Harvard University Press 2005). Nussbaum's response is quite different from mine. She thinks the instability, and resulting inadequacy of extant forms of contractarianism requires their supplementation with non-contractarian ideas deriving from an Aristotelian version of virtue ethics, broadly understood. My approach, first pursued in the first edition of this book, is to identify a stable, and I argue viable, form of Kantian contractarianism purged of all offending (and unnecessary) Hobbesian elements.
8. David Gauthier, *Morals by Agreement* (Oxford: Oxford University Press 1986).
9. John Rawls, *A Theory of Justice* (Oxford: Oxford University Press 1971); *Political Liberalism* (Oxford: Oxford University Press 1993).
10. Rawls, *A Theory of Justice*, pp. 7–11.
11. *Ibid.*, pp. 302–3.
12. This point is made with admirable clarity by Will Kymlicka, *Contemporary Political Philosophy* (Oxford: Oxford University Press 1990), p. 25.
13. Rawls, *A Theory of Justice*, pp. 100–8.
14. *Ibid.*, p. 12.
15. Michael Sandel, *Liberalism and the Limits of Justice* (Cambridge: Cambridge University Press 1982).
16. Rawls, *A Theory of Justice*, p. 138.
17. See especially, 'Justice as fairness: Political not metaphysical', *Philosophy and Public Affairs* 14, 3 (1985): 225–51.
18. It is this rather crucial point that seems to be continually overlooked by communitarian critics of Rawls.
19. See, for example, R. M. Hare, 'Rawls theory of justice', in N. Daniels ed., *Reading Rawls* (New York: Basic Books 1975). Brian Barry, *The Liberal Theory of Justice* (Oxford: Oxford University Press 1973).

20. Rawls, *A Theory of Justice*, p. 121.
21. *Ibid.*, p. 20.
22. This point applies with particular obviousness to Carruthers's version of contractarian theory expressed in *The Animals Issue*.
23. The analogy is borrowed from Simon Caney, 'Liberalism and communitarianism: A misconceived debate', *Political Studies* 40, 2 (1992): 277.
24. See his *Problems of Philosophy* (Oxford: Oxford University Press 1912), ch. 5.
25. Rawls, *A Theory of Justice*, p. 508.
26. *Ibid.*, p. 131.
27. *Ibid.*, p. 505.
28. *Ibid.*, pp. 505–6. Emphasis is mine.
29. *Ibid.*, p. 504.
30. *Ibid.*, p. 505.
31. *Ibid.*, p. 512.
32. *Ibid.*, p. 20.
33. This is the claim of the necessity of origin, defended by Saul Kripke, *Naming and Necessity* (Cambridge, MA: Harvard University Press 1980).
34. I say no worry, but this is tendentious. Those concerned with the environment, for example, might want to extend the scope of morality in precisely the way that I have argued contractarianism rules out. I think that, in fact, contractarian approaches can yield a substantive environmental ethic. It is just that the moral status they will accord the non-animate world will necessarily be an indirect one.
35. Carruthers *The Animals Issue*, p. 102.
36. Peter Singer, *The Expanding Circle: Ethics and Sociobiology* (Oxford: Oxford University Press 1981).

7 Animal minds

1. The distinction is originally due to Bertrand Russell.
2. See Stephen Stich, *From Folk Psychology to Cognitive Science* (Cambridge, MA: MIT Press 1983), p. 5.

3. See my 'Consciousness and higher-order thoughts', in *Mind and Language*, 16, 3 (2001): 290-310, and also *The Nature of Consciousness* (Cambridge: Cambridge University Press 2001), ch. 5.
4. Peter Carruthers *Language, Thought and Consciousness* (Cambridge: Cambridge University Press 1998).
5. Norman Malcolm 'Thoughtless brutes', *Proceedings of the American Philosophical Society* 46, (1973): 5-20.
6. Donald Davidson, 'Rational animals', in E. LePore and B. McLaughlin eds, *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (Oxford: Basil Blackwell 1985), pp. 473-80. See also Davidson's 'Thought and talk', in S. Guttenplan ed., *Mind and Language* (Oxford: Oxford University Press 1975), pp. 7-23.
7. Davidson, 'Rational animals', p. 475.
8. *Ibid.*, p. 475.
9. See Davidson, 'Mental events', in his *Essays on Actions and Events* (Oxford: Oxford University Press 1980), pp. 207-25.
10. Content holism has played a central role in Davidson's writings in the philosophy of language. See his *Inquiries into Truth and Interpretation* (Oxford: Oxford University Press 1984).
11. Stephen Stich, 'Do animals have beliefs?', *Australasian Journal of Philosophy* 57, (1979): pp. 15-28.
12. Stich, *From Folk Psychology to Cognitive Science*, pp. 104-6.
13. Tom Regan, *The Case for Animal Rights* (London: Routledge 1988), pp. 49-60.
14. Davidson, 'Rational animals', p. 475.
15. David DeGrazia, *Taking Animals Seriously: Mental Life and Moral Status* (New York: Cambridge University Press 1996), I believe deGrazia falls victim to the temptation described here.
16. Hilary Putnam 'The meaning of "meaning"', in K. Gunderson ed., *Language, Mind and Knowledge: Minnesota Studies in the Philosophy of Science* 7 (Minneapolis: University of Minnesota Press 1975). Tyler Burge 'Individualism and the mental', *Midwest Studies in Philosophy* 4 (1979). See also Burge's 'Individualism and psychology', *Philosophical Review* 95 (1986): 3-45.
17. Colin McGinn 'The structure of content', in A. Woodfield ed., *Thought and Object* (Oxford: Oxford University Press 1982), pp. 207-58.

18. Davidson, 'Rational animals', p. 475.
19. Ibid., 475.
20. Ibid., 476.
21. Ibid., 477.
22. Ibid., 479.
23. Ibid., 479.
24. See, for example, Jerry Fodor and Ernest LePore, *Holism: A Shopper's Guide* (Oxford: Basil Blackwell 1991).
25. Ruth Millikan, *Language, Thought and Other Biological Categories* (Cambridge, MA: MIT Press 1984).

Bibliography

- Anscombe, Elizabeth (1958) 'Modern moral philosophy', *Philosophy* 33: 1–19.
- Barry, Brian (1973) *The Liberal Theory of Justice* (Oxford: Oxford University Press).
- Burge, Tyler (1979) 'Individualism and the mental', *Midwest Studies in Philosophy* 4.
- Burge, Tyler (1986) 'Individualism and psychology', *Philosophical Review* 95: 3–45.
- Caney, Simon (1992) 'Liberalism and communitarianism: A misconceived debate', *Political Studies* 40, 2.
- Carruthers, Peter (1992) *The Animals Issue: Moral Theory In Practice* (Cambridge: Cambridge University Press).
- Carruthers, Peter (1996) *Language, Thought and Consciousness* (Cambridge: Cambridge University Press).
- Carruthers, Peter (1998) 'Natural theories of consciousness', *European Journal of Philosophy* 6: 53–78.
- Clark, Stephen (1977) *The Moral Status of Animals* (Oxford: Oxford University Press).
- Davidson, Donald (1970) 'Mental events', in his *Essays on Actions and Events* (Oxford: Oxford University Press), pp. 207–25.
- Davidson, Donald (1975) 'Thought and talk', in S. Guttenplan ed., *Mind and Language* (Oxford: Oxford University Press), pp. 7–23.
- Davidson, Donald (1984) *Inquiries into Truth and Interpretation* (Oxford: Oxford University Press).
- Davidson, Donald (1985) 'Rational animals', in E. LePore and B. McLaughlin eds, *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (Oxford: Basil Blackwell), pp. 473–80.
- DeGrazia, David (1996) *Taking Animals Seriously* (Cambridge: Cambridge University Press).

- Dretske, Fred (1995) *Naturalizing the Mind* (Cambridge, MA: MIT Press).
- Fodor, Jerry and LePore, Ernest (1991) *Holism: A Shopper's Guide* (Oxford: Basil Blackwell).
- Gauthier, David (1986) *Morals by Agreement* (Oxford: Oxford University Press).
- Gibson, James (1979) *The Ecological Approach to Visual Perception* (Boston, MA: Houghton-Mifflin).
- Griffin, J. (1986) *Well Being: Its Meaning, Measurement and Moral Importance* (Oxford: Oxford University Press).
- Hare, R. M. (1975) 'Rawls theory of justice', in N. Daniels ed., *Reading Rawls* (New York: Basic Books).
- Hare, R. M. (1984) 'Rights, utility and universalization: A reply to J. L. Mackie', in R. G. Frey ed., *Utility and Rights* (Minneapolis: University of Minnesota Press).
- Harsanyi, J. (1976) *Essays on Ethics, Social Behaviour and Scientific Explanation* (Dordrecht: D. Reidel).
- Hursthouse, Rosalind (1999) *On Virtue Ethics* (Oxford: Oxford University Press).
- Hursthouse, Rosalind (2000) *Ethics, Humans and Other Animals* (New York: Routledge).
- Kant, Immanuel (1963) 'Duties to animals and spirits', in his *Lectures on Ethics*, trans. L. Infield (New York: Harper & Row).
- Kripke, Saul (1980) *Naming and Necessity* (Cambridge, MA: Harvard University Press).
- Kundera, Milan (1983) *L'Insoutenable Légèreté de L'Être* (Paris: Gallimard).
- Kymlicka, Will (1989) 'Contractarianism', in P. Singer ed., *A Companion to Ethics* (Oxford: Basil Blackwell).
- Kymlicka, Will (1990) *Contemporary Political Philosophy* (Oxford: Oxford University Press).
- Malcolm, Norman (1973) 'Thoughtless brutes', *Proceedings of the American Philosophical Society* 46: 5-20.
- McGinn, Colin (1982) 'The structure of content', in A. Woodfield ed., *Thought and Object* (Oxford: Oxford University Press), pp. 207-58.
- McGinn, Colin (1992) *Moral Literacy, or How to do the Right Thing* (Cambridge, MA: Hackett).
- Midgeley, Mary (1984) *Animals and Why They Matter* (Harmondsworth: Penguin).

- Mill, J. S. (1968) *Utilitarianism, Liberty, Representative Government*, ed. A. D. Lindsay (London: J. M. Dent & Sons).
- Millikan, Ruth (1984) *Language, Thought and Other Biological Categories* (Cambridge, MA: MIT Press).
- Moore, G. E. (1912) *Ethics* (Oxford: Oxford University Press).
- Nussbaum, Martha (2005) *Frontiers of Justice* (Cambridge, MA: Harvard University Press).
- Putnam, Hilary (1975) 'The meaning of "meaning"', in K. Gunderson ed., *Language, Mind and Knowledge: Minnesota Studies in the Philosophy of Science* 7 (Minneapolis: University of Minnesota Press).
- Rachels, James (1990) *Created From Animals: The Moral Implications of Darwinism* (Oxford: Oxford University Press).
- Rawls, John (1971) *A Theory of Justice* (Oxford: Oxford University Press).
- Rawls, John (1985) 'Justice as fairness: Political not metaphysical', *Philosophy and Public Affairs* 14, 3: 225–51.
- Rawls, John (1993) *Political Liberalism* (Oxford: Oxford University Press).
- Regan, Tom (1984/1988) *The Case for Animal Rights* (London: Routledge).
- Regan, Tom (1985) 'The case for animal rights', in Singer ed., *Taking Animals Seriously*.
- Rosenthal, David (1986) 'Two concepts of consciousness', *Philosophical Studies* 49: 329–59.
- Rosenthal, David (1990) 'A theory of consciousness', Report 40/1990 Center for Interdisciplinary Research (ZiF), Research Group on Mind and Brain, University of Bielefeld.
- Rosenthal, David (1993) 'Thinking that one thinks', in G. Humphreys and M. Davies eds, *Consciousness* (Oxford: Basil Blackwell).
- Rowlands, Mark (1997) 'Contractarianism and animal rights', *Journal of Applied Philosophy* 14, 3.
- Rowlands, Mark (2001) 'Consciousness and higher-order thoughts', *Mind and Language* 16, 3: 290–310.
- Rowlands, Mark (2001) *The Nature of Consciousness* (Cambridge: Cambridge University Press).
- Rowlands, Mark (2002) *Animals Like Us* (London: Verso).
- Rowlands, Mark (2008) *The Philosopher and the Wolf* (London: Granta).
- Russell, Bertrand (1912) *Problems of Philosophy* (Oxford: Oxford University Press).
- Sandel, Michael (1982) *Liberalism and the Limits of Justice* (Cambridge: Cambridge University Press).

- Sapontzis, Stephen F. (1987) *Morals, Reasons, and Animals* (Philadelphia: Temple University Press).
- Scruton, Roger (1996) *Animal Rights and Wrongs* (London: Demos).
- Singer, Peter (1974) 'All animals are equal', *Philosophical Exchange* 1, 5.
- Singer, Peter (1975/1991) *Animal Liberation* (London: Thorsons).
- Singer, Peter (1979) 'Killing humans and killing animals', *Inquiry* 22.
- Singer, Peter (1980) *Practical Ethics* (Cambridge: Cambridge University Press).
- Singer, Peter (1980) 'Animals and the value of life', in *Matters of Life and Death*, ed. T. Regan (New York: Random House 1980).
- Singer, Peter (1980) 'Utilitarianism and vegetarianism', *Philosophy and Public Affairs* 9, 4.
- Singer, Peter (1981) *The Expanding Circle: Ethics and Sociobiology* (Oxford: Oxford University Press).
- Singer, Peter (1985) *Taking Animals Seriously* (Oxford: Basil Blackwell).
- Slote, Michael (2001) *Morals From Motives* (Oxford: Oxford University Press).
- Smart, J. J. C. (1973) 'An outline of a system of utilitarian ethics', in J. Smart and B. Williams eds, *Utilitarianism: For and Against* (Cambridge: Cambridge University Press).
- Stich Stephen (1979), 'Do animals have beliefs?', *Australasian Journal of Philosophy* 57: 15–28.
- Stich, Stephen (1983) *From Folk Psychology to Cognitive Science* (Cambridge, MA: MIT Press).

Index

- Aggregation requirement, 42, 46,
47, 48, 49, 51
- Anscombe, Elizabeth, 98
- Aquinas, 120
- Aristotle, 98
- blood sports, 101, 102, 103,
104, 105
- Burge, Tyler, 198
- Carruthers, Peter, 3, 121, 160, 161,
162, 181, 182, 183, 186
- communitarianism, 136, 137, 142
- Confucianism, 98
- consciousness, 6, 181, 182, 183,
184, 185, 186, 187
- contractarianism, 2, 3, 4, 5, 118,
119, 120, 121, 123, 124, 125,
126, 127, 128, 129, 130, 131,
132, 133, 134, 135, 136, 137,
138, 139, 140, 141, 142, 143,
144, 145, 146, 147, 148, 152,
153, 154, 155, 156, 157, 162,
163, 168, 171, 172, 173,
174, 175, 176, 180
- Hobbesian vs. Kantian, 123,
124, 125, 126, 127, 128, 130,
131, 147, 148, 152, 153, 155,
156, 157, 158, 174, 175
- contractualism, *see*
contractarianism
- Davidson, Donald, 6, 187, 188, 192,
194, 195, 196, 197, 203, 204,
206, 207, 208, 209, 210, 211,
212, 218
- de dicto* vs. *de re* (attributions of
belief), 201, 202, 203, 204,
205, 206, 207, 208
- deGrazia, David, 196
- Descartes, René, 218
- Dretske, Fred, 181
- Dualism, 88, 89
- equal consideration, principle
of, 29, 31, 32, 42, 43, 46, 47,
48, 49, 50, 51
- equality of power condition, 124,
125, 127, 148, 156
- experience, 178
- experience machine, 33, 34
- Gauthier, David, 130
- genetic fallacy, 161
- Gibson, James, 208
- Griffin, J., 41
- Hare, R. M., 41, 42

- harm principle, 67, 68, 69,
80, 81, 82
- Harsanyi, J., 41
- higher-order thought (model of
consciousness), 6, 181, 182,
183, 184, 185, 186, 187
- Hobbes, Thomas, 4
- holism of the mental, 187, 188, 189
- Hursthouse, Rosalind, 100, 101,
102, 103, 104, 105, 106, 107,
108, 109, 110, 112, 113
- husbandry
animal, 53, 54, 55, 56, 57, 79, 80,
81, 83, 165, 166
human, 165, 166
- Independence Day scenario, 9,
11, 26
- intuitions
reflective, 65
unreflective, 140, 142, 154
- intuitive equality argument, 133,
134, 138, 139, 141, 148, 150,
151, 152, 154, 175
- Iroquois, 29
- Kant, Immanuel, 5, 120, 127, 128
- Kripke, Saul, 159
- Kundera, Milan, 113, 117
- Kymlicka, Will, 5, 220n
- liberty principle, 79, 80, 81, 82,
84, 85
- Maher, Bill, 111
- Malcolm, Norman, 187
- marginal cases, argument
from, 22, 23, 28
- mercy, virtue of, 113, 114,
115, 116, 117
- Midgeley, Mary, 108
- Mill, J. S., 41
- Millikan, Ruth, 213
- minimize harm principle, 77
- miniride principle, 71, 72, 73, 74,
76, 77, 78, 81
- Moore, G. E., 38
- moral considerability, 17, 20
- moral rights 1, 2, 3, 59, 68, 69,
70, 71, 72, 73, 76, 77, 78,
118, 119, 120
as valid claims 69, 70, 71
direct 119, 120
indirect 119, 120
natural 2
- Morlocks (and Eloi), 167, 168
- Namuhs, 9, 10, 11, 12, 13, 14, 15, 16,
17, 18, 19, 20, 21, 22, 23, 24, 30
- naturalistic fallacy, 63, 64
- opaque versus Transparent
(attributions of belief), 204,
205, 206, 207, 208, 217
- original position, 5, 132, 135, 136,
137, 142, 143, 144, 145, 147, 151,
157, 160, 162, 173, 175
metempsychotic version of, 146,
147, 155, 156
- Plato, 98
- propositional attitudes, 176, 177,
178, 179, 180, 187, 188, 189, 190,
191, 192, 193, 202, 203, 204,
205, 206, 207, 208, 209, 210,
211, 213, 214, 215, 216, 217

- Putnam, Hilary, 198, 199, 200, 202
- Quine, Willard van Orman, 202
- rationality condition, 124, 125,
127, 148, 157
- Rawls, John, 5, 37, 38, 50, 65, 97,
118, 121, 122, 128, 131, 132,
133, 134, 135, 136, 137, 138,
139, 140, 141, 142, 143, 144,
147, 149, 151, 152, 153, 154,
155, 156, 157, 159, 161, 174
- reference, 197, 198, 199, 200
- reflective equilibrium, 140
- Regan, Tom, 1, 2, 3, 6, 38, 54, 55,
58, 59, 60, 61, 62, 63, 64, 65,
66, 67, 68, 69, 70, 71, 72, 74, 75,
76, 77, 78, 81, 82, 83, 84, 85, 87,
88, 89, 90, 92, 93, 94, 95, 111,
112, 118, 121, 176, 179, 180,
193, 194
- relational proper functions, 213,
214
- respect principle, 66, 67, 69, 72, 75,
76, 78, 79, 81
- Rosenthal, David, 181, 182
- Russell, Bertrand, 145
- Scruton, Roger, 100, 101, 102,
103, 104, 105, 106, 107,
108, 112, 113
- sensations, 176, 177, 178, 180
- Singer, Peter, 1, 2, 31, 32, 41, 53, 54,
55, 161, 162, 179, 180
- Smart, Jack, 33, 34, 35
- social contract argument, 135,
136, 137, 138, 139, 150, 175
- speciesism, 27, 28
- Stich, Stephen, 6, 187, 189, 190,
191, 192, 194, 195, 196, 197,
199, 203, 204, 207, 212, 218
- subject-of-a-life, 58, 59, 60, 61, 63,
68, 70, 86, 87, 90
- supervenience, 14, 87
- teleological theory of
representation, 213, 214,
215, 216
- theory of mind, 182
- Trollope, Anthony, 108
- twin earth, 198
- utilitarianism, 1, 6, 32, 33, 34, 35,
36, 37, 38, 39, 40, 41, 42, 43, 44,
45, 46, 47, 48, 49, 50, 51, 52, 53,
54, 55, 56, 57, 62, 65, 66, 89, 90,
91, 92, 176, 179
- and animals, 53–7
- act vs. rule, 51, 52, 53
- egalitarian version of, 39, 40,
41, 42
- hedonistic, 33, 34
- preference, 34, 35, 35, 36
- teleological version of, 37, 38,
39, 41
- utility, definition of, 32–6
- value
inherent, 1, 38, 58, 59, 61, 62, 63,
64, 65, 66, 70, 72, 74, 75, 78, 84,
86, 87, 88, 89, 90, 92, 93, 94, 95,
96, 97
- intrinsic, 62, 63
- and valuing, 93, 94, 95, 96, 97
- veil of ignorance, 132, 135, 136,
137, 143, 146, 147, 150, 158, 166

virtues

concept of, 98, 99, 113
conflicts of, 101, 103, 104
list of, 100
moral vs. executive, 111, 112
subjectivity of, 101, 103
vagueness of, 101, 103

virtue ethics, 2, 3, 6, 97, 98ff

Wells, H. G., 167

Wittgenstein, Ludwig, 61, 207

worse-off principle, 74, 75, 76, 77,
78, 81, 83, 86