

OXFORD

A painting of a small sailboat on a river. The boat is light-colored with a large white sail. A person in a dark jacket and cap is seated in the boat. The water is calm with a reflection of the boat. The background shows a riverbank with trees and buildings under a pale sky.

ON HUMAN RIGHTS

James Griffin

6

The Metaphysics of Human Rights

6.1 TWO MODELS OF VALUE JUDGEMENT

There is a ‘taste model’ of value judgement, given its classic statement by Hume, still highly influential in philosophy, all but dominant in the social sciences, and now widely absorbed into common sense.¹ According to it, value judgements are a matter of taste or attitude: you have your opinion; I have mine. Each of our opinions can be corrected for factual or logical error, but once that is done, there is no further ground for regarding one value judgement as better than another. Factual judgements can be true or false; value judgements are neither. Factual judgements are objective; value judgements are subjective—subjective in both of the two most common senses. They are, first of all, merely expressions of taste or attitude. And, second, values are not part of the furniture of the world; the world contains physical objects, properties, events, minds, but it does not also contain values. When philosophers thought they were deriving natural rights from human nature—that is, from empirical facts about human beings—they were really deriving them from human interests, using the word ‘interest’ to mean what is in one’s interest or to one’s advantage. That is, they were deriving human rights from value judgements that other cultures might not share.

The taste model is obviously appealing, as attested by its current popularity. But it seems to me to collapse, in the end, from its own explanatory inadequacy. According to the taste model, our preference fixes on an object, which thereby becomes valuable. But value cannot be explained so simply. There is no reliable correlation between our *actual* preferences and what is *valuable* to us. It is a discouraging, and not uncommon, fact of life that one can get what one actually wants, even sometimes what one most wants in life, only to find that one is no better off.

So we might, as many philosophers and social scientists do, drop *actual* preferences in favour of *rational* preferences. We can accept, as surely we

must, that a thing does not become valuable just by being desired; our desires can be based on false or incomplete information. We must understand more fully or more accurately what the natural world is like, and only then, in this enhanced state of knowledge, might our reactions of desire be directed at what can count as a value. This 'rational preference account' is much the more plausible, and among philosophers now the more common form of the taste model. But what standard does 'rational' represent? We might say, as Richard Brandt proposes, that a desire is rational if it persists when I have become aware of all the relevant natural facts and when I have purged my thought of logical error.² But is this enough? Take an instructive example that we owe to John Rawls.³ A man has a particularly crazy aim in life—say, counting the blades of grass in various lawns. He knows that no one is interested in the results, that the information is of no use, and so on; he commits no logical error. But we should be hard put to it to see the fulfilment of this obsessive desire as enhancing his life—apart, that is, from preventing anxieties and tensions that might be set up by frustrating his desire, but that is to introduce other values. What we should be hard put to see is the fulfilment of his desire as, in itself, improving the quality of his life. Or, if Rawls's example is thought doubtful because the man does not exhibit normal human rationality, take a woman with a sadistic streak who, after years of psychotherapy, knows full well how much her sadistic behaviour harms herself but finds it too intensely pleasurable to give it up. As she is a normally reasonable person, she will most likely have two desires: to give up her sadism and not to give it up. The first is for her the most sensible desire; the second is the motivationally more powerful desire. But both desires pass the test set by the taste model: they survive confrontation with all relevant facts and logic. The mere persistence of a desire does not make its object good for us. The first desire is the rational one; her sadism does her more harm than good. But that is a form of rationality that the taste model has trouble accommodating.

What these examples suggest is that our standard for 'rational' has not become strong enough yet. The way to make it stronger, though, is to make desires 'rational' in some such sense as 'formed in appropriate appreciation of the nature of their object'. But though this seems to handle the counter-examples, it seems also to undermine the preference account of value. It stresses an *appropriate* reaction of desire, and so suggests that there is an element here of getting things right. Once the idea of the 'appropriateness' of a response enters, standards of correctness and incorrectness enter. The taste

model has no ready answer to the question, When is a response appropriate? The mere fact that the term 'rational preference' retains the word 'preference' does not show that much of the taste model is surviving. One cannot answer that the appropriate response is to be understood as the 'natural' or 'normal' one. If 'normal' here is taken to introduce some sort of statistical standard such as 'most common', then we may well find that most of us, even when informed, go on wanting certain things—say, to assert ourselves—too much; it may just be an unfortunate tendency in human nature that we have to struggle to keep in check. If 'normal' is taken to mean something closer to 'correct', then that is just the stronger standard that we are trying to explain.

We have more critical resources than the taste model recognizes. We can ask more searching questions about our aims, and resort to more radical criticism in answering them, than the taste model allows.

First, there is what has to be in place for language even to be possible. A word has meaning only in virtue of there being rules for its use, rules that settle whether or not the word is used correctly. Wittgenstein has argued that these rules cannot, in the end, be satisfactorily understood as a template that we carry in our heads—an image, say, or a list of defining properties—but only as part of shared practices in a community. And these shared practices are possible only because of the human beliefs, interests, dispositions, sense of importance, and so on that go to make up what he called 'a form of life'.⁴ Our form of life provides the setting in which our language develops and only within which it is intelligible. And a form of life seems to consist in part in a shared set of beliefs and values. It is because, and only because, we see others as, like us, understanding the world in a certain way, caring about certain things, regarding certain things as important, that we can communicate with one another. So, among other things, shared values are needed: such things as that we feel pain and ordinarily dislike it and want to avoid it or have it alleviated, that we aim at certain kinds of things and can be gratified or frustrated. Donald Davidson has a similar argument.⁵ We cannot, he thinks, interpret the language that others use without assuming that we have certain beliefs and attitudes in common with them—that, for instance, many of our aims, interests, desires, and concerns are the same. If that is right, then general scepticism about basic common-sense values is self-defeating. Certain values are part of the necessary conditions for our language, which sets for us the bounds of intelligibility. These arguments of Wittgenstein and Davidson seem to me persuasive; the difficulty is to say how far they take us. How many

such basic beliefs are there? They will be confined, I think, to a few of the most basic human interests, the interests that I have already mentioned: that we want to avoid pain and anxiety, that we have goals and attach importance to their being fulfilled (and perhaps also a few moral norms closely connected to these interests, such as that cruelty is wrong). But in any case, what Wittgenstein and Davidson say is a start in finding the various sources of our beliefs about human interests, still using 'interests' in the sense I explained a short while ago.

Then there are interests outside the central core needed for intelligibility. Suppose that one day I am struck by the thought that your life seems, in some way, better than mine. I have a sense that I am frittering away my life in trivia, and you strike me as accomplishing things with your life that give it weight or point. My thought might initially be quite ill-focused, and I should have to try to sharpen it. Not just any achievement of yours would contribute much weight—say, walking the length of Broad Street in Oxford on your hands. It will have to be the achievement of something that is itself valuable. But that is not enough either. Some values are just too small-scale to give one's life weight or point. I should have to go on in this vein, trying to isolate what it is that I think so valuable in your life. I should be driven to use value-rich vocabulary to bring this possible value into focus: 'accomplishment' (if I may simply commandeer this word for what I am after) is roughly the sort of achievement that gives life weight or point. Then, having isolated it, having distinguished it from other values and from the valueless, I should have to decide whether what is left really is valuable—or rather, as the search for the definition already brings in value-rich language, these two processes, definition of the possible value and decision about its value, go hand in hand. And one decides about its value not by appeal to one's own subjective set of desires. There is nothing there to appeal to, except the vacuous desire to have a good life, which does not do the job, because the present job is to decide specifically whether accomplishment, so defined, makes a life good. What seems to be playing a big role here is not my subjective set of desires, but my understanding what accomplishment is. This sort of understanding, which has its own standards of success, might therefore introduce a new value into my life, in a way that the taste model fails to explain. Call this second model 'the perception model'.

For me to see anything as enhancing my life, I must see it as enhancing life in a generally intelligible way, in a way that pertains to *human* life and not just to my particular life. Why should this be so? It runs counter to

widespread belief. One reason why we resist this conclusion is that we tend to overlook the constraints that are part of the constitution of desires: desires of the sort we are interested in—that is, the ones that have links to values—are not just brute psychological responses to objects that we can delineate in purely natural terms. To think that they are is to overlook the fact that we must also see the object as in some way good. There are, of course, different sorts of desires. Some desires are, in effect, afflictions: for instance, cravings, obsessions, compulsions, post-hypnotic suggestions, addictions, habits. We passively observe their occurrence in us. But there are also desires that are part of normal intentional action, part of the sort of behaviour that makes up the vast bulk of human life. We have options; we reflect, choose, and act—none of it necessarily very consciously. Desires of this sort aim at the good. An agent's normal behaviour is to recognize interests and to act to meet them. This sort of desire is the sort that concerns us in an account of values, and it fails on its own terms if it does not aim at something that we take to be good. It essentially involves a judgement of good, even if only a primitive form of one.

Another reason why we resist this conclusion is that it seems to fly in the face of the plain fact that people are very different from one another and get very different things out of life. I may greatly value playing the piano well; you may value expert rock climbing; and neither of us may care a fig for what the other cares about. But merely caring about something does not make it valuable. For anyone to see anything as valuable, from any point of view, requires being able to see it as worth wanting. This is a perfectly general requirement on values; it is the basis of the distinction between mere wanting and the sort of wanting that connects with values. One way to see something as worth wanting is to see it under the heading of some general human interest. Anyone who thinks that not all values are like that must then explain in what further way we can see them as worth wanting. What could make playing the piano well something worth wanting is that it would be in some way rewarding: I should enjoy it, or it would be an accomplishment, and so on. And you will have to do the same with your rock climbing. To see anything as making life better, we must see it as an instance of something generally intelligible as valuable and, furthermore, as valuable for any normal human being.

Deliberation about human interests ends up, I think, with a list of values. I am less concerned with precisely what is on the list than I am with the conclusion that deliberation ends with a general profile of values, a

chart of the various high points that human life can rise to. My own list (no doubt incomplete) is this: accomplishment, enjoyment, deep personal relations, certain kinds of understanding, and—the interests that are most immediately relevant to human rights—the components of personhood. I shall return to the last item on the list shortly.

6.2 HUMAN INTERESTS AND THE NATURAL WORLD

Is ‘human nature’, as we use the term in accounts of human rights, part of the natural world? Can claims about human interests be correct and incorrect, true and false, in the way that claims of fact can be?

Claims about human interests based on biological needs can be correct and incorrect. One plain human interest connected closely to pain and survival is in nourishment: without it we suffer or die. Another plain interest is in certain kinds of human contact: if a baby is fed but denied other forms of nurture, it will suffer great psychological damage and might die. There are clear criteria for judgements that nourishment and other forms of nurture are human interests: namely, that they avoid egregious disvalues: ailment, pain, and malfunction. These plain disvalues are part of the framework necessary for the intelligibility of language.

Now look at values not in that framework. One of them would be the example I used earlier, accomplishment. If I accomplish nothing in my life, I shall suffer; my life will lack point or weight. But it is an altogether less experiential sort of suffering than the gross ailment and malfunction of a baby deprived of nurturing.

Still, these values outside the framework of intelligibility are both continuous with the values in it and, like them, firmly embedded in human nature. Particularly deeply embedded in us are certain biological aims—for food, health, protection of our capabilities—and certain psycho-biological aims—for example, for companionship, affection, reproduction. But we are not only intentional animals; we are also reflective animals. It does not take us much reflection to see that goods differ in degree, and that many are good only as means, often remote means, to other things that are good as ends. We come to see that the goods we aim at day in and day out are mostly trivial or mere means. It is natural for intentional, reflective beings to form second-order desires. We want something more than the satisfaction of trivial wants or wants for mere means. We want the whole activity, the unstoppable

succession of desire and fulfilment, to be itself sometimes leading to what is neither trivial nor a mere means. That is, I think, the characteristic aim of reflective, intentional beings. It is characteristic even though often not an especially conscious aim. But we form a desire to get out of the 'rat race', or we wonder whether we would not be better off sacrificing some income for a better quality of life. And it takes only a whiff of our own mortality for just about any of us, reflective types or not, to wonder whether we have wasted our lives. These large-scale, course-of-life desires emerged at the time of the evolutionary transition from a merely biological being to a reflective, intentional one. Non-biological interests, such as accomplishment, are as deeply embedded in human nature as biological ones are. To put it briefly, we are rational animals; biological interests are embedded in our animal nature, and non-biological ones in our rational nature.

The shift from biological to non-biological interests brings with it other changes that we should note: for instance, the move from predominantly experiential forms of harm such as pain and ailment, which are fairly easily identified, to non-experiential sorts of harm. How do I know that lack of food is harmful? Well, obvious physical symptoms appear. How do I know that lack of accomplishment is harmful? Well, because life is empty in a certain way—namely (and here circularity threatens), it lacks meaning or point.

Are there criteria for those non-experiential sorts of harm that would allow judgements about them to be correct or incorrect? The only plausible account of how we identify something that is in our interest, such as accomplishment, gives a role to both recognition and reaction, but without separating them nearly as sharply as the taste model does. The taste model says that value judgement involves, first, recognizing certain features of the natural world and then, second, reacting to them with approval or disapproval. But, as we saw, to explain value, we need not mere reaction but *appropriate* reaction. And to explain the appropriateness of a reaction, we need more than just a description of the object in purely natural or factual terms—that is, where the words 'natural' and 'factual' have the boundaries given them by seventeenth-century science or eighteenth-century philosophy, say, in the form of David Hume's sharp distinction between 'fact' and 'value'. For example, we bring what I am calling 'accomplishment' into focus only by resorting to such terms as 'fulfils life', and such language is not value-neutral but already organizes our experience by selecting what we see favourably. Being 'fulfilled', in the sense needed, is not a psychological matter of having a feeling of fulfilment; it is

a matter of life's not being empty or futile or wasted. Of course, a notion such as 'wasted' is itself evaluative, so a value is already built into our notion of 'life-fulfilling', a value that makes itself felt in not just any object's being a possible object of fulfilment. Approval is not left free to fix on one object or another; its direction is already fixed in, and manifested by, what we see favourably. Reaction here is nothing as simple as a sentiment of approval; certain standards of appropriateness are essential to its being the reaction that it is. And recognition is not itself fully describable without the introduction of some reactive elements.

There is Aristotle's question, Are things valuable because desired, or desired because valuable?⁶ The taste model answers: the first. The perception model, on certain well-established interpretations, answers: the second. But there is a third answer: neither; there is no priority. I want to recommend the third answer. I shall regard it as a variation on the perception model, but one could, if one preferred, give it a name of its own.

So the notion of recognition is best understood as something not entirely reaction-free, and vice versa. 'Recognition', in this appropriately impure sense, is meant to be a kind of sensitivity to something in the world. But one is not entitled to talk in terms of a sensitivity unless one can explain what it is for the sensitivity to work well and what to work badly. We have a clear, well-established account of the working of the physical senses. A theory of what goes on in the world includes an explanation of what goes on in human perception. We rely on our perceptions to justify the theory, and the theory to justify our reliance on our perceptions. But the theory explains, among other things, why perceptions in certain conditions are as a body reliable, how they fail, and how we can sometimes detect and correct their failure. Is anything approaching this possible for value judgements? Unless it is, why should we think that we have got a sensitivity *to* anything in the world?

Now, one might see hope of developing an account of the notions of 'correct' and 'incorrect' applied to *moral* judgements if one thought (as I do) that these judgements are grounded in some way, not necessarily a consequentialist way, in (largely human) interests. But we are now concerned with judgements about human interests themselves, and they seem, as value judgements go, to be ground-floor. Still, although ground-floor, the sensitivity I mean is complex in its workings and rich in its connections. We can say a fair amount about what it is for it to work well. One needs, first of all, a lot of knowledge of the familiar, undisputed factual sort about the world.

One has also to have sufficient human capacities to know how enjoyment, say, figures in human life. In this way, one can build up an account of the conditions for the successful workings of our sensitivity to prudential values, akin to conditions such as good light, good eyes, and good position for successful seeing. The account of failure in the sensitivity is independent of most judgements that the sensitivity should deliver. To show that it failed, one would have to show that the person concerned lacked the concept, or information, or certain human capacities, and the test for lacking any of them is fairly well removed from the deliverances of the supposed sensitivity. For instance, to show that I lack certain capacities for feeling, you would have to go to empirical psychology or to biology to show how the difference from a normal human psyche came about. One ought to be able to build up an account of the conditions in which, if all are met, the sensitivity succeeds. If so, this sensitivity will differ from a sensitivity such as sight, not in there being no account of its working, but in the greater difficulty of knowing when the conditions for its working are met. It is not that we never get evidence that they are all met. A full account of deliberating about human interests suggests that in the right conditions we are sensitive to certain things' making life go better. There are, of course, other explanations of what is going on besides the existence of such a sensitivity, but there are examples that make them seem implausible. For instance, we aim at some things simply because of deep, largely invisible social pressures. But there are also persons who come up with new (to them, at least) value notions, such as accomplishment, that have never been taught to them and the ethos of whose society is live-for-the-moment. In the end the best explanation of such changes is that the person has hit upon, has become sensitive to, something valuable, and that its being valuable is to some extent independent of the process of coming to regard it as such. A sensitivity to values would also be a good explanation of convergence of belief between persons, especially if the convergence emerged when the conditions of reliability were present, but most else—social ethos, psychological bent, economic class—were all very different. One could hope for simpler, more direct evidence, but this evidence is neither out of reach nor negligible.

So I think that we may conclude that judgements about human interests can be correct or incorrect. They report deliverances of a sensitivity to certain things going on in the world: namely, interests being met or not met. These interests are part of *human* nature, and not just human nature as seen by society. These judgements seem to be correct or incorrect, not, say, in the

way that conclusions in mathematics can be, but rather true or false in the way that statements of natural fact can be.

The notion of ‘meeting an interest’ is rather like the notion ‘soothes’: something is relieved. We think of something’s being soothing and something’s meeting an interest as both being, to put it in rough, intuitive language, not properties *in* objects but properties *of* objects, properties that relate to the objects’ interaction with other things. Of course, there is a rich, well-understood causal base for a judgement about something’s being soothing. If we want some enlightenment about the authority that our ethical beliefs have, we must know somewhat more about the kind of truth they have—for instance, whether we may regard statements about human interests as statements of natural fact.

Well, a statement about being soothing and a statement about meeting interests must be much like one another because, on closer look, the first statement *is* an instance of the second. An ointment, say, soothes an irritation, and an irritation is in the general class of pains and discomforts, which are cases of disvalues. Compare ‘That ointment soothes my irritation’ with ‘That accomplishment makes my life fulfilled’. In the second judgement, too, a value enters to explain why people are in certain respects as they are—namely, with interests met or unmet. It explains why some people suffer from a sense of emptiness or futility, especially at the end of life, whereas others do not. The value can be at work on us even without our being conscious of it—even, indeed, without our having the concept of ‘accomplishment’. The absence of the value can explain the vague, unfocused dissatisfaction with life that can come before we are able to explain it.

Now, advocates of the taste model will, of course, resist my current line of thought. One does not have to cite a *value* (as if one were talking about something in the world), they are likely to reply, in order to explain this sense of emptiness; all that one needs to cite is a *belief* that one’s life is empty, and all that one needs to cite to explain the vague, unfocused sense of emptiness is a vague, unfocused belief. But this reply falls short at two points. First, it goes no way towards explaining why the emptiness in question occupies much the same sort of place in our life as does an irritation that some ointment might soothe. Both are lacks that are part of human nature. It also ignores where the belief that one’s life is empty itself comes from. The best explanation of why so many people form the deathbed belief that their life has been empty may well be that there has been a characteristic human interest often unmet.

6.3 THE TEST OF THE BEST EXPLANATION

I suggested earlier that it seems best to think in terms of our possessing a sensitivity to certain values in which recognition and reaction are merged—that is, the perception model, so adumbrated. We recognize a lack, an interest; furthermore, we recognize that certain things fill the lack or meet the interest. That is, we recognize a value by recognizing certain things that characteristically go on in human life. The best explanation of certain people's belief that something or other is to our advantage is that there are features of human life that they recognize. The best explanation of, say, someone's coming up with a new (to him or her) value notion such as accomplishment is that the person has become sensitive in this way to an interest to be met, to a value.

Then there is, as I mentioned briefly earlier, the phenomenon of convergence of beliefs between several persons. Most facts about convergence and divergence in normative beliefs are neutral as between the taste model and the perception model. What matters to the choice between these two views, however, is what *explains* the convergence or divergence. This is a complicated empirical issue. But if, when certain knowledge and sensitivity and conceptual equipment are all in place, convergence in belief occurs, and if that happens when other causal influences on the formation of belief, such as social pressures, are different, then the best explanation may turn out to be the workings of the recognition that I was just referring to. Certainly, if the explanation I suggested earlier in the case of one person's coming to recognize the value of accomplishment is plausible, it will be a likely candidate in the many-person case.

I have been employing the test of the best explanation. We conceive of a belief-independent world of empirical fact as playing a role in affecting our concepts, our beliefs, and our sometimes converging on the same beliefs. We attribute existence in the world of empirical fact to a kind of thing—any kind of thing, including values—in assigning a certain sort of explanatory role to it. One sort of explanatory role that would seem to have these existential implications—sufficient for them, but perhaps not necessary—is causal. This gives us a test of empirical existence in terms of the best causal explanation: a kind of thing has empirical existence if things of that kind must appear in the best account of what happens in the empirical world. If

entities such as electrons or properties such as electrical charge must appear in the best account of what happens in the world, then there are such entities and such properties. The same would be true of human interests and of events such as these interests' being met or unmet.

The test of the best explanation is often taken to be a test of realism about a thing. What realism claims, including realism about values, is very difficult to settle. Much of language has meaning only within the context of natural human concerns, desires, interests, sense of importance, and so on—Wittgenstein's idea of a 'form of life'. This raises the possibility that the embeddedness of our concepts in the human point of view is so deep and inescapable that it makes no sense to speak of a belief-independent reality. There may be a case, for all that I have said, for a wide irrealism that would carry a narrower irrealism about values along with it. But I have not here used the test of the best explanation as a test of the all-things-considered reality of values, but only of their factuality. What I have been after is some fuller understanding of the way in which judgements about human interests can be true, and I am content to conclude no more than that they can be true in the way that judgements about 'soothing' can be.

So my proposal is this. A judgement about some accomplishment's being life-fulfilling and a judgement about an ointment's being soothing are both judgements about what goes on in the world of (human) nature. They are true in virtue of that part of the natural world. That the concepts 'soothing' and 'life-fulfilling' are deeply embedded in the human perspective, that recognizing their occurrence necessarily involves a human response, that the world of (human) nature may not be entirely mind-independent, does not undermine the possibility of their having a truth-value. A typical human response goes into making something 'soothing' and also 'life-fulfilling', rather than being (as the taste model represents it) a truth-destroying part of the criteria for judging it actually to be soothing or life-fulfilling. And nature consists of objects, properties, and events that are independent of our ideas and beliefs about them in the following sense. Our ideas are shaped by what they are ideas of; we can alter an idea, or even drop some ideas and invent new ones, as we discover more about nature or just reflect more deeply on what we already know. And we confirm our beliefs against nature—that is the truistic version of the correspondence theory of truth. We look more closely; we collect evidence; we find counter-examples. These are the ordinary ways in which we established the truth of a claim that a certain ointment soothes a certain irritation.

What I propose is that whether certain human interests are met is also a matter of fact. It is *also* a matter of value. So my suggestions run contrary to the sharp separation of fact and value present in the taste model.

That I am in pain is a matter of fact; to that everyone agrees. Now, pains have both a phenomenological side to them (the internal feel of our experiences of pain), and, equally important, an active side (reactions of avoidance, alleviation, and so on). One learns the word 'pain' both by having certain experiences and by understanding where pain fits into human life—that 'pains' are characteristically (though not, necessarily universally) to be avoided or alleviated. This reactive element cannot be sharply separated from the recognitional element. And this is so not because, with pain, recognitional elements and reactive elements, though separable in principle, are difficult to disentangle, but because the distinction between these two kinds of elements ceases to hold here. And that is because our standard of sameness in the sensations that we bring together under the concept 'pain' is partly that they are characteristically what are to be avoided, alleviated, and so on. With pains, we do not recognize something to which we also, independently, react. Our reaction is a constituent of our recognition. The way in which pain fits into human life is part of the criterion for its being pain.⁷ And it is not that, as a matter of fact, we just find ourselves desiring to avoid pain or to have it alleviated. What is going on is more complex: we have these desires *because* we find pain undesirable. We have an attitude towards it; we find it bad, and for obvious reasons. Some basic values are part of the framework necessary for language, and the disvalue of pain must be one of them. The distinction between fact and value, as we find it in Hume and more generally in the taste model, becomes difficult to sustain at this point.

Cases of interests' being met and unmet, I think, earn their way into the world of facts. We can place them in our everyday natural world, and do not need to resort to anything remotely like a detached 'value realm'. That conclusion brings out what seems to me immensely plausible about ethical naturalism. In talking about human interests we are not talking about entities in such an other-worldly realm—detectable, say, by intuition—but, rather, about certain things that happen in the only realm that values need: mainly, what goes on in human lives, that *this* or *that* meets an interest, and so makes a life go better. It makes sense to ask how these sorts of happenings relate to other, fairly well-defined levels of explanation: say, the psychological. But I doubt that there is any point in asking how such happenings relate to a level

so grossly defined as the ‘natural’ or the ‘empirical’ or the ‘factual’, because the boundaries that we use to delineate the ‘natural’ or the ‘empirical’ or the ‘factual’ are not just fuzzy, but so central to what we must settle as to make assumptions about where they are located question-begging. We do not start our investigations with these boundaries satisfactorily drawn. We have only a common, extremely vague intuition about the ‘natural’ or ‘empirical’ world, one that is full of contentious ontological assumptions. So we should not start by asking how values relate to ‘fact’, as if we really knew the territory inhabited by ‘fact’ and were wondering only about its foreign relations. That procedure makes their relation more puzzling than it needs to be. I suggest that our notion of the ‘factual’ is wide enough to include events of meeting and failing to meet interests. The right position, therefore, is, after all, a kind of naturalism. But it is not the usual kind of reductive naturalism, in which the boundaries of the ‘natural’ or the ‘factual’ are kept relatively tight: that is, roughly in the position that they have long had in the fact/value split in Hume’s view that values cannot be derived from facts. What seems attractive, however, is an expansive naturalism, in which the boundaries of the ‘natural’ or the ‘factual’ are pushed outward a bit, in a duly motivated way, with the effect that they now encompass human interests.

6.4 THE METAPHYSICS OF HUMAN RIGHTS

How much of the epistemic and metaphysical standing of human interests carries over to the human rights derived from them? In turning now from human interests to human rights, we move from prudence to morality.

Here is an example of a relatively easy such move. ‘That’s cruel’ is a judgement about action, but is short of commanding action, as ‘ought’ and ‘must’ judgements do; so it avoids some of the further complexities that arise with them. All the same, ‘That’s cruel’ already encapsulates a standard of behaviour. How does this moral standard arise? The full answer to that question will draw on evolutionary biology, anthropology, psychology, decision theory, as well as on the patterns of justification that especially concern ethics. My partial answer will draw primarily on the last.

A person who acts cruelly intends to make another suffer without compensating good. That intention is both necessary and sufficient. If I do something just to hurt you (say, twist your arm) but, by a fluke, save you greater pain (say, by replacing your dislocated shoulder), what I do is none the less cruel.

If I try to help you (say, by replacing your dislocated shoulder) but cause you greater pain (say, by breaking your arm), I am clumsy or oafish but not cruel.

But, then, is ‘That’s cruel’ not, after all, a moral judgement, but a factual one about intent? John Mackie thinks so. ‘What is the connection’, he asks, ‘between the natural fact that an action is a piece of deliberate cruelty—say, causing pain just for fun—and the moral fact that it is wrong?’⁸ Mackie accepts the taste model: we delineate the action through the factual description ‘cruel’ and then respond to it with disapproval. But the notion ‘cruel’, I think, leaves no space for disapproval to be an independent step. To understand ‘pain’ already involves regarding it as a disvalue. We typically respond to pain negatively, and its having that standing in our life often gives me a reason to avoid causing you pain, and so to avoid being cruel to you.⁹

There are prudential values and disvalues so basic, so centrally embedded in our conceptual framework—pain, for instance—that the idea of deliberation to reach the conclusion that it is a value or a disvalue does not fit the case. One cannot make sense of those looming presences in our life, other people, without understanding how, with their vulnerable bodies and psyches, they fit into the world—not least, their being able to be hurt by it. Deliberation about the value of pain is left no space to get going. When we move on to moral deliberation, we find much the same thing. A moral notion such as ‘cruel’, being conceptually so close to ‘pain’, inherits much of its obviousness. How would one establish that I had a reason not to be cruel to you? My reason comes partly from inevitable features of our conceptual framework: my seeing you as a person involves my accepting that there are certain basic values at stake in your life, and my seeing them as values produces a reason for me to respect them. Again, deliberation has no space to get going.

The obviousness of the judgement ‘That’s cruel’ comes also from its generally being made well within certain motivational limits. It is well within the capacities of the human will not to torture cats for fun; the most ordinary people manage it. And it costs us nothing, at least nearly all of us, not to torture cats for fun. The judgement ‘That’s cruel’ generally operates within an area in which the human frame can easily manage the required action, so the condemnation built into the word ‘cruel’ is apt.

The same is true of many norms. Sometimes the standard for behaviour they set is well within human capacity. If we are willing to make the judgement ‘That’s cruel’, with its condemnatory force, we should be willing to accept the norm ‘Don’t be cruel’.

The moral judgement ‘That’s cruel’ does not go much beyond claims about pain and, importantly, intention. The property ‘cruel’, being a combination of pain, causes, and intentions, has whatever metaphysical standing they have—standing as natural facts, I should say.

Some philosophers would insist that ‘cruel’ has prescriptive force, which my analysis solely in terms of natural facts about intention, cause, and pain fails to capture. One should concede to the objectors that ‘cruel’ does indeed have prescriptive force, if that means merely that it gives a reason for action. The explanation of its reason-giving status, I have suggested, is roughly the same as the one for ‘pain’. ‘It hurts’ can generate a reason for me to avoid the thing myself and also a reason for me not to visit it on you. The reason-generating force of ‘cruel’ derives from, and has the same explanation as, the reason-generating force of ‘pain’. There is no residual force that needs further explanation of a natural-fact-undermining sort.

But let us now turn to an example of the second kind of transition from prudence to morality: not smooth, hardly irresistible, and perhaps involving considerable change in epistemic and metaphysical standing. Recall the earlier discussion of the norm ‘Don’t deliberately kill the innocent’.¹⁰ Human life is of especially high value. From that we derive an especially strict norm, ‘Don’t deliberately kill the innocent’. Part of the content of the right to life—indeed, in the seventeenth century much of it—is not killing another person without due process, though as far back as Locke, philosophers were including certain further protections—for example, aiding the desperately needy.¹¹ But, as I said earlier, when we make the move from human interests to moral norms, certain limitations of human agents enter the picture, especially limitations in will and understanding.¹² There are limits to what may be demanded by human agents, and those limits help to shape the content of the human right to life—for example, the matter of how much aid for the needy may be demanded of one. And the limits of understanding limit what is available to us to decide the content of human rights. Sometimes we can calculate reliably enough—that is, to a degree of probability on which we should be prepared to act—the consequences of large-scale, long-term social arrangements. We can, if the changes in question are extreme enough. But in less extreme cases, often in just the cases we regard as live options, we cannot—cases such as deliberately killing non-combatants in war, using terrorism as a political instrument, killing one patient to save five others, and so on. Here too we often cannot do the calculation of consequences to a reliable degree of probability. And there is no obvious remedy for this failure.

In the absence of a remedy, we find some other way to conduct our moral life. Or rather, in many cases a sufficiently reliable, all-encompassing calculation of consequences has never been available to us, so we have simply carried on with our moral life pretty much in the piecemeal, not fully systematic way that mankind has always done. We have at times raised our standards and broadened the considerations that concern us, but we have not convincingly risen to an overarching system. Instead, we long ago developed a different approach to ethical decision making. The very great value of human life has led to our having very great respect for it; we allow that there can be exceptions to the norm 'Don't deliberately kill the innocent', but out of our great respect for human life we demand that the case for any exception be especially convincing. That is, we respect life: we do not try to promote it, for example, by maximizing it. There is an element of policy in this approach, and we can see that another society might adopt a somewhat different policy. As this norm constitutes a large part of the content of the human right to life, there is therefore an element of policy in the human right as well.

The policies I refer to are not, at least typically, consciously chosen. It is not that we become aware of our limitations of will and understanding and then, in conjunction with others, decide in light of these limitations to adopt a certain policy. We do not discover our limitations, because it never seemed to us that our will and understanding were unlimited. Instead, such policies usually emerge in a society without anything so deliberate as a group decision. Nor are these policies arbitrary; they are shaped in large part by, and are largely a response to, the great value that we attach to human life. And the policies can be criticized. They can be too strict or too lenient, too demanding or too undemanding, too limited as to object or too unlimited, and so on. For example, not long ago the policy in our society that 'charity begins at home' was often interpreted to allow charity to end there too; but now we think that our duties of help can sometimes be worldwide. Our actual, present-day ethical policies may, no doubt, be inadequate, but we are able to improve them. Because of our limitations of understanding and will, moral philosophy cannot realistically aspire ultimately to abolish this element of policy; its more realistic, but still ambitious, aim is to arrive at the best policy.

This element of policy is not peculiar to the right to life. One of the two grounds for human rights is practicalities.¹³ The content of most human rights becomes sufficiently determinate only by considering certain practicalities. The personhood ground tells us, for example, that we have a right to

security of person; without it there would be no security of agency. But the personhood ground, on its own, does not yield a line nearly determinate enough to tell us what in practice is prohibited. For that we need to consider human psychology and the ways in which societies function, and decide whether we need a safety margin, and roughly how generous it should be. And here too an element of policy enters. The line we have plumped for may not be quite the same as the line another society has plumped for—and not necessarily because of a difference in our societies but because of a difference in the policies that our two societies have happened to plump for.

If this element of policy is a necessary determinant of the content of many human rights, what does this mean for their metaphysical standing? The norm ‘Don’t deliberately kill the innocent’ expresses a policy. An expression of a policy is not true or false in the way that a statement that a human interest is or is not met is; it is not a matter of natural fact. Of course, one difference is that a moral policy is often expressed in imperative mood. Still, even if we were to translate the norm ‘Don’t deliberately kill the innocent’ from imperative to declarative mood, and to express it in a form more amenable to assessment in terms of truth and falsehood—say, ‘Deliberately killing the innocent is wrong’—nothing important would have changed. The best understanding of the latter form of words would still be, in part, an utterance of a policy, and policies attract assessment not in terms of truth and falsity, but in terms of how well they perform their function. That might tempt one to think that the statement ‘Deliberately killing the innocent is wrong’ reduces to the claim ‘The policy of prohibiting deliberately killing of the innocent is a good one’, and that statement may in turn be reducible to the claim that the policy performs its function well—as the claim ‘That is a good pen’ may be reduced to the claim that it performs the functions of a pen well, which *may* itself be a statement of natural fact. But, of course, one should not be tempted. Expressing a policy is not assessing it. We are not asserting that it is a good policy, though we should not adopt a policy if we knew that it were on balance a bad one, or that there were a substantially better one. But at a fairly early point in assessing policies such as ‘Don’t deliberately kill the innocent’, we reach a point where we can no longer tell that one policy is better than another.

7

The Relativity and Ethnocentricity of Human Rights

7.1 ETHICAL RELATIVITY

Ethical relativism, as I shall understand it, makes two claims: first, that ethical judgements are made within a framework of basic evaluations, which may take the form of beliefs, preferences, sentiments, and so on; and, second, that there are divergent frameworks for judgements on the same matter, no one framework being most authoritative.¹ We can then specify the framework further case by case—the basic evaluations of individual persons, of social groups, of cultures, and so on.

Ethical relativism, as most commonly expressed, is universal: *all* ethical judgements are relative to a framework. Its contradictory is therefore particular negative: some are not. I argued in the last chapter that some ethical judgements—namely, judgements about basic human interests—are objective, where ‘objective’ means dependent not upon a person’s subjective states but upon considerations that would lead all successfully rational persons to the same conclusion. So universal ethical relativism is, I conclude, false. But philosophers tend to treat values as if they were uniform: all are objective, or none is; all are a matter of knowledge, or none is; all are relative to a framework, or none is. But in the last chapter I also questioned this assumption of uniformity. Some complex moral norms, such as ‘Don’t deliberately kill the innocent’, have an element of policy to them, and so lack empirical truth-value, whereas the judgements that a particular human interest is or is not met have one.

Relativism need not take a universal form. Consider relativity to the evaluative framework of individuals, based on their different desires and sentiments. Value beliefs can be subjected to criticism by facts² and by logic.³ Many ethical beliefs are shaped by a person’s understanding, often misunderstanding, of the empirical world: of the consequences of our acts,

of what the objects of our desires are really like, and so on. Once one's desires and attitudes have been corrected, one may come to change them; over time they may increasingly converge with the desires and attitudes of others. What a relativist must maintain, however, is that some divergent beliefs will remain, and remain for the reasons relativists give. In any case, our interest here is human rights. Are *they* relative to a framework?

How can one make a case for ethical relativism? The commonest way is to cite, with little in the way of argument, certain examples of particularly stubborn ethical disagreement, which are meant to leave one thinking that the best explanation of the disagreement is the relativist's. This is, of course, an extremely weak form of argument. Establishing the best explanation of stubborn ethical disagreements requires understanding all the possible origins of these conflicting beliefs and all the possible resources that might resolve the conflict—no quick or easy job. That the job is so difficult leaves many relativists, despite its inadequacy, doing no more than citing examples. Let me give a brief sampler of the examples that they have offered.

Some societies regard theft as a serious crime; others do not even have the concept of private property, on which the idea of 'theft' depends.⁴ It is hardly obvious that relativism provides the best explanation of this difference. If one lives where food is plentiful without cultivation, there may be no pressure to develop an institution of private property. But if one's survival depends upon clearing land and shouldering the burdens of growing one's own food, some form of control over the land and the crop is highly likely to emerge. The best explanation may be difference not in ethical framework but in material conditions.

Some societies have tolerated infanticide; others condemn it.⁵ But consider the extreme case of life-threatening poverty. Tolerance of infanticide is an adaptation that most of us would make if forced to it by the direst poverty: say, if one were faced with the awful choice between the survival of one's newborn baby or one's young child. A plausible explanation of the disagreement over infanticide between a society of such abject poverty and one better off may not be a difference in evaluative frameworks but, again, a difference in material conditions.

Many people are committed to preserving the environment; others see no objection to exploiting it.⁶ This is a conflict that does indeed look irresolvable. To my mind, we can coherently talk about the value of the environment not

just when changes in the environment affect human beings, say our health or enjoyment, but also apart from any effect on sentient life. The environment has a value in itself. The idea of the environment's being intrinsically valuable rests, I believe, on an idea of appropriateness of attitude. The only appropriate response to, say, the enormous age, biological complexity, and beauty of the Great Barrier Reef is wonder and awe. And wonder and awe prompt respect. There is something lacking in a person who does not have some such response. The wanton destruction of the Great Barrier Reef would be a monstrous act. Ethics, I should say, is broad enough to encompass standards not just of *right* and *wrong* but also of *appropriate* and *inappropriate*. Now, if the natives on an island in the Great Barrier Reef decide to improve their quality of life by mining, and thereby destroying, the Reef, the apparent rational resolution of the conflict between the preservationists and the exploiters would be to weigh the costs and benefits to sentient creatures against the intrinsic value of the Reef. But that, I suspect, is a piece of weighing we cannot do. We must remember that some values may be incommensurable, in this sense of the term: two values are incommensurable if and only if they cannot be ranked against one another as 'greater than', 'less than', 'equal to', or 'roughly equal to'.⁷ For a pair of values to be commensurable in this sense, there must be a bridging notion in terms of which the comparison between them can be made. For example, most, perhaps all, human interests, I should say, lend themselves to comparison. They do, not because there is a substantive super-value behind them, but because there is a formal value notion in terms of which we can, and regularly do, compare them: for example, 'prudential value', 'quality of life', or 'human interest' itself. We thus have the conceptual materials to judge that 'this would enhance the quality of my life more than that', 'this is a more major human interest than that', and so on. But sometimes—not often, I believe—two competing values are so different in nature from one another that there is no bridging notion available. In this conflict over the environment, for example, there is no bridging notion; comparison breaks down. This is indeed an intractable difference, but it does not derive from difference in ethical framework but from incommensurably different values. There is even a possible resolution of this disagreement: bringing both parties to see that the values they purport to commensurate are incommensurable.

A last example. Many of us think that abortion is prohibited; many others think that it is permitted.⁸ Most often a person who holds that abortion is forbidden also holds background religious beliefs. But then is this, after all,

an example of ethical relativity? Virtually all of us would accept that abortion is prohibited if we believed that an all-good, all-wise God had told us so. But with such a background, this intractable disagreement seems to have arisen not from different ethical frameworks but from different metaphysical beliefs. Perhaps, though, this just means that we should reconsider our definition of ethical relativism as relativity to a framework of basic *evaluations*. Evaluations cannot be sharply divided from empirical and metaphysical beliefs; our basic evaluations are what they are in part because of non-ethical beliefs. But if this truth is to support the relativity of a belief about the morality of abortion, it must be because of the further relativity of facts or of metaphysical conceptual schemes. Ethical relativity would then not stand alone. Although these further relativities seem much shakier than ethical relativity, perhaps that impression is mistaken.

Still, not all ethical disagreements about abortion arise from differences over religion. When they do not, what best explains the stubbornness of the divergence? No doubt, many different things. But one explanation that is hard to make plausible is that there are two different frameworks of fairly well-articulated and well-defined ethical beliefs producing this disagreement. That would make thought at this level far clearer and more inferential than it is. What might these ethical beliefs be? Nor is it plausible that these divergent beliefs about abortion are themselves basic ethical beliefs. They do not have quite that depth; they need justification themselves. What is more plausible, I should say, is that the framework for each of these conflicting views is a complex mix of ethical beliefs, factual beliefs, and sentiments. They might be beliefs such as ‘A foetus is already a fully biologically formed potential person, as much so as a new-born baby’ or ‘An early foetus is too biologically primitive to be a person’. Or they might be sentiments such as revulsion at the very thought of killing a foetus or, on the contrary, equanimity in the face of it. But these beliefs are vague, and their implications for action by no means clear. And we should have to decide what weight to attach to these sentiments of revulsion or equanimity. What authority do they have?

My discussion of each of the four examples I have given is, I admit, inconclusive—neither decisively for nor against their relativity. But that is my point. One would have to dig much deeper before one could reach a satisfactory conclusion. Merely citing an example is no case at all. Let me now try to dig somewhat deeper in the example that primarily concerns us: human rights.

7.2 THE RELATIVITY OF HUMAN RIGHTS

Human rights are suspected—by Westerners as much as by Easterners—of being relative to Western culture. Human rights are undoubtedly a Western product: introduced by Christians in the late Middle Ages and further developed there in the early modern period and in the seventeenth and eighteenth centuries.⁹ They were part of the growth in individualism in that particular time and place; they were part of a new sense in Europe and the Americas of ‘the dignity of man’ and the great value of human autonomy and liberty.

But why think that human rights are, as well as a product of the West, also relative to the values of the West? One argument might be that the values from which human rights are derived—most prominently autonomy and liberty—are themselves peculiarly Western values. Some societies, it is true, value autonomy highly, seeing in it the peculiar dignity of the human person, while other societies value autonomy much less, seeing in it the threat of social atomism and the loss of solidarity and fraternity and of the harmony that comes from our all serving the same values. But anyone who thinks seriously about the value of our status as normative agents and the benefits of living in a cohesive fraternal community will recognize that both are highly important. And they will recognize the same about both others’ having to respect our individuality and our having duties of concern and care for others. It may be that realizing certain of the values of individualism is incompatible with realizing certain of the values of community. But incompatibility of values is not their relativity. Besides, the frequency of the incompatibility is exaggerated. Not all forms of autonomy are the autonomy to which we attach great value.¹⁰ I would display more autonomy, in one correct use of the word, if I calculated my own income tax each year and decided for myself the plausibility of the Big Bang, instead of relying on the expertise of others. But neither of those is the autonomy to which we attach great value. What we attach great value to is the autonomy that is a constituent of normative agency, and relying on a tax accountant or an astrophysicist does not derogate in the least from one’s normative agency. And the form of solidarity to which we attach such great value does not require surrendering our normative agency, though it may require greater trust in one another and greater convergence in public standards. The form of solidarity that is of great value is a joint commitment to the members of one’s community

and to the community's successful working. The plausible explanation of the fact that different societies rank autonomy and solidarity differently is not that they are rankings of the relativist sort. Everyone, on pain of mistake, has to admit that autonomy and solidarity are both highly valuable. No one would maintain that any loss in autonomy is worse than any loss in solidarity, or vice versa. And the more specific a choice between the two becomes—a certain loss of autonomy, say, to achieve a certain gain in solidarity—the more convergence in choice one will expect there to be. We do seem able, if only roughly, to compare these competing values.

A second argument for the relativity of human rights—indeed, an argument arising from my own account—is this. In the last chapter we saw how certain moral judgements—for example, 'That's cruel'—could be derived from judgements about human interests—for example, 'That's painful'.¹¹ The judgement 'That's cruel' goes so little beyond claims about pain, cause, and intention that it inherits the metaphysical and epistemic standing they have—standing as natural facts, I proposed. This suggests—merely suggests—that a human right (a moral standard) might similarly be derived from a certain human interest (a prudential value), again inheriting from it a sort of objectivity that would defeat the claim of relativity. Take the derivation of autonomy (the human right) from autonomy (the prudential value). But I also admitted that the derivation of still other human rights from human interests was less simple—for example, the right to life, which has an element of policy to it. The norm 'Don't deliberately kill the innocent', which is one of the correlative duties of the right to life, in part expresses a policy, and different societies might adopt different policies. Some human rights thus have a clear conventional element. Do they thereby have an element of relativity?

Take the right to autonomy. Once one recognizes the value of autonomy, one recognizes also a reason to be autonomous oneself and a reason not to deny other people their autonomy. Human rights are protections of one's personhood, and so protections of, among other things, one's capacity for and exercise of autonomy. Is the objective epistemic status of the judgement that autonomy is prudentially valuable transferred to the judgement that autonomy is a human right? We should ask: What more comes into the second judgement than is already present in the first? The obvious answer is: the first is a prudential judgement, the second a moral judgement. I find it very hard to understand the nature of the transition from prudence to morality, but, despite my uncertainty, I think that at least there is a kind of

rationality to it. It is tempting to treat the reason-generating consideration that moves me when my autonomy is at stake as different from the one that moves me when yours is at stake. The obvious difference between these two cases is that in the one it is *my* autonomy, and in the other it is *yours*. But the most plausible understanding of the engine of these two judgements is *autonomy: because a person's quality of life is importantly at stake*. The *my* and *your* are not part of the reason-generating consideration. The clause *because a person's quality of life is importantly at stake* lacks reference to me or to you, but it lacks nothing of what we understand the reason to be. To try to deny 'autonomy' its status as a reason for action unless it is attached to 'my' would mean giving up our grasp on how 'autonomy' works as a reason for action.

Return now to my question: What more is present in the second judgement than is already contained in the first? There is, of course, whatever is added by calling autonomy a 'human right'. Many philosophers say that the judgement that something is a human right carries with it a claim that it has a particular moral importance: for example, it has the status of a 'trump' or a 'side-constraint'. But I have already argued several times against this characterization of human rights. They are neither trumps nor side-constraints. They are not even the most important of rights. Autonomy—or, more generally, personhood—is not necessarily the most important human interest. Human rights make only an overrideable claim that a person's autonomy be given due respect—that is, the respect due to the sort of autonomy at stake in any particular case. And that much follows simply from autonomy's being a prudential value. It is true that to know that autonomy is a prudential value is also to know how valuable it is: that it is generally highly valuable to us, valuable enough to attract, as it has, special protection, but of varying value from case to case, and overrideable by other important values.

When I speak here of the 'derivation' of the human right to autonomy, I do not mean an entailment. I mean only that a reasonable person who recognizes the prudential value of autonomy will also recognize the respect that it is due. And the reasonableness of that transition is enough to deny a relativist a foothold here.

Another important qualification. The transition from prudence to morality is, of course, too complicated a matter to be dealt with as briskly as I have just done—so complicated that there is no point in my embarking on a few more brisk comments. I have discussed the subject more fully elsewhere,